认知科学导论

[加] P.萨伽德 著
朱菁 译

中国科学技术大学出版社

责任编辑：田天恩
封面设计：邹云霞

作者保罗·萨伽德现为加拿大滑铁卢大学哲学系教授，心理系和计算机科学系兼职教授。

译者朱菁分别于1991年和1994年获中国科学技术大学工学学士（计算机科学技术）和理学硕士（科学技术哲学）学位，现为中国科学技术大学研究生院（北京）讲师。

认知科学之所以成为一门成熟的基础科学，是因为它已形成了其特有的科学的基本概念和方法论。因此像学习任何一门成熟的基础科学一样，要了解认知科学，就必须系统地学习它的基本概念和方法论，及其理论和假设。在这样的意义上，萨伽德所著《认知科学导论》一书中译本的出版为想真正学习和了解认知科学的读者提供了一本难得的中文入门教材。

——陈霖　中国科学技术大学教授，中国科学技术大学北京认知科学开放研究实验室主任，中国科学院－北京医院脑认知成像中心主任。



ISBN 7-312-00939-5/Q·19
定价：10.00元

认知科学导论

P·萨伽德 著

朱 菁 译

中国科学技术大学出版社

1999·合肥

内 容 简 介

认知科学是对心智与智能的跨学科的研究，包括心理学、人工智能、神经科学、语言学、人类学和哲学。本书以表征-计算这一看待心智的核心观念为线索，系统地介绍和评价了当代认知科学中的不同研究路线所取得的成就，包括基于逻辑、规则、概念、类比、表象和联接（神经网络）所发展的各种认知科学理论。同时也讨论了从情绪、意识、物质环境与社会环境、动力学系统和数学知识等方面对以表征-计算观为代表的正统认知科学的挑战。本书是为各种不同学科的大学生和低年级研究生撰写的教材，内容精练，深入浅出，力求把握当前认知科学发展的最新动向。对有关领域的教师与研究人员亦很有参考价值。作者保罗·萨伽德是加拿大滑铁卢大学哲学系教授，心理学系和计算机科学系的兼职教授。

图书在版编目（CIP）数据

认知科学导论/（加拿大）萨伽德（Thagard，P.）著；朱菁译.－－合肥：中国科学技术大学出版社，1999.5

书名原文：Mind：Introduction to Coqnitive Science

ISBN 7-312-00939-5

Ⅰ．认…　Ⅱ．①萨…②朱…　Ⅲ．认知科学-概论

Ⅳ．B842.1

中国版本图书馆CIP数据核字（1999）第16522号

中国科学技术大学出版社出版发行

（安徽省合肥市金寨路96号，邮编：230026）

肥西新华印刷厂印刷

全国新华书店经销

850×1168·32　印张：6.75　字数：173千

1999年5月第1版　1999年5月第1次印刷

印数1－－4000册　定价：10.00元

Mind: itroduction to cognitive science

著作权合同登记号：皖登字 1299025 号

前　言

认知科学是对心智(mind)和智能(intelligence)的跨学科的研究,包括哲学、心理学、人工智能、神经科学、语言学和人类学。它的学术起源在50年代中期,不同领域的研究者开始借助复杂的表征(representation)和计算程序来发展关于心智的理论。到70年代中期,其学术组织开始形成,成立了"认知科学学会"并开始出版《认知科学》杂志。至今,在北美和欧洲,已有60多所大学设立了认知科学专业,更多的学校则开设了认知科学的课程。

讲授认知科学这门跨学科课程有相当的难度,因为学生来自各种不同的背景。从1993年起,我在滑铁卢大学开设了名为"认知科学导论"的课程。一方面,这门课吸引了来自计算机科学和工程等学科的学生,他们具有较好的计算方面的背景却对心理学和哲学所知甚少;另一方面,在心理学和哲学上有良好背景的学生却对计算不甚了解。这本教材便是为建设一门毋须任何认知科学的专业准备的课程所作尝试的一个部分,其目的是使对心智和智能感兴趣的同学能够认识到在对心智的研究中有许多互补的途径。

向来自不同学科的读者介绍认知科学至少有三种方式。第一种是集中介绍各个不同的领域,如心理学、人工智能,等等;第二种是以心智的不同功能来组织讨论,如问题解决、记忆、学习和语言。我选择了第三种方式,即系

统地介绍和评价认知科学家们提出的主要的关于心理表征(mental representation)的理论,包括逻辑的、规则的、概念的、类比的、表象的和联接的(神经网络)理论。对这些基本理论途径的讨论是为了以一种统一的方式来展示认知科学不同领域所取得的成就,同时注重对各种重要的心理功能的理解。

撰写此书的目的是为所有有兴趣选修认知科学导论课的学生提供一本入门书籍。为了达到这一目的,就需要以一种适当的方式向心理学系的学生介绍逻辑,向英语系的学生介绍计算机算法,向计算机科学系的学生介绍哲学上的争论。作者热情地欢迎各方面的读者提出宝贵意见。

虽然本书是针对本科生的,对于研究生和教师也有参考价值,有助于他们了解他们各自的领域在整个认知科学的事业中所处的状况。我没有试图写成一部百科全书。由于整个着眼点是提供一个整体性的导论,我既要使得全书的篇幅保持在适当的范围内,同时也力求侧重于森林而不是树木。因为将认知科学视为各个不同领域的交叉融合而不是一个统一的整体,我忽略了人工智能、认知心理学和心智哲学等导论课程里的许多标准论题。在每一章的最后都附有小结和进一步的推荐读物。每章的"备注"一节为有特别兴趣的读者提供信息。

本书强调了心理表征理论对理解心智的重要性,同时也意识到认知科学还有很长的路要走。第二部分讨论了对认知科学基本假设的挑战,并且对这一跨学科研究未来的方向提出了建议。

目　　次

第一章 表征与计算

探索心智

不知你是否曾经对于你的心智是如何工作的感到过惊奇？人们每天都要完成各式各样的心理任务：在工作和学习中解决问题，对个人生活作出决定，对所知道的人的行为给予解释，以及获取各种新的概念。认知科学的主要目的就是要去解释人们是怎样完成这各种各样的思维活动的。我们不仅要对各种问题求解和学习的过程进行描述，还要对心智是怎样去执行这些操作的给出说明。

理解心智如何工作对于许多实践活动来说都至关重要。教育工作者需要了解学生思维活动的本质，以便寻求更好的方法来进行教学。工程师和其他设计人员则需要知道他们的用户在有效或无效地使用他们设计的产品时是怎么想的。通过反思是什么使人变得聪明，我们可以使得计算机变得更具智能。而政治家和决策者们如果能理解与他们打交道的人们的心理过程，他们将会取得更大的成功。

然而对心智的探索实非易事，因为我们不可能剥开一颗心智来看一看它究竟是如何工作的。千百年来，哲学家和心理学家将心智作了各式各样的比喻，例如将它比作一张白纸，在上面做各种印记；或是一种液压装置，由各种力来操作它；甚至将它比作电话开关板。近半个世纪来，更新的、富于启发性的比喻来自于各种新型计算机的发展。许多（但不是全部）认知科学家都将思维视作一种计算，并采用计算隐喻来描述和解释人们是怎样学习和求解问题的。

你所知道的是什么？

在同学们进入大学后，除了课程内容之外还有许多东西是需要了解的。不同专业的学生将要学习的课业千差万别，但对于大学是如何运作的都需要有所了解。如何选课？课程什么时间开始？哪些课程好而哪些课要尽量避开？要获得学位必须具备哪些条件？从一栋教学楼到另一栋哪条路径最优？校园里的其他同学是怎样的？哪儿是欢度周末的最佳场所？

这些问题的答案会成为大多数学生脑海中的一部分，但这是怎样的一部分呢？大多数的认知科学家都认为心智中的知识是由心理表征（mental representation）构成的。大家都熟悉各种非心理表征，比如说印在本页上的词语。在这里我就用了“本页”一词来代表你正在阅读的这一页。同学们会经常使用到一些形象化的表征，例如校园和教学楼的地图。为了说明像学生对大学的了解这些不同类型的知识，认知科学家们提出了各种各样的心理表征方式，如规则、概念、表象（image）和类比。学生要了解的规则有“如果我要能毕业，那么我必须选够10门本专业的课程”。他们会增加一些新的概念，例如像使用“小鸟”、“米老鼠”或是“香肠”这样一些词汇来形容一门特别容易的课。为了从一栋楼房转到另一栋楼，一幅校园布局的心理表象或图片将是十分有用的。在选修了一门特别喜欢的课程之后，可能还会选择一门类似的课程来学习。通过与校园里不同专业的同学接触，心目中可能会形成关于某种类型的学生的模板(stereotype)，尽管很难准确地说出究竟是什么构成了这样一些模板。

学生掌握这样一些有关校园生活的知识并非只是为了积累信息。作为学生要面对很多的问题，诸如如何学好各自的专业，怎样拥有合宜的社会生活，以及怎样在毕业后找到工作。解决这些问题就需要与心理表征打交道，例如作出“为了能够毕业还必须

再修 5 门课”这样的推断，或是作出“千万别再选梯迭姆（Tedium）教授的课”这样的决定。认知科学提出人们是通过在心理表征之上运行心理程序（mental procedure）而产生出思维和行动，而规则和概念这些不同类型的心理表征则支持不同类型的心理程序。在此我们不妨看一下表示数目的不同方式。绝大多数人熟悉的是阿拉伯数字系统，1，2，3，10，100，等等，以及进行加、乘等运算的标准程序。罗马数字同样可以表示数目，用Ⅰ、Ⅱ、Ⅲ、Ⅹ、C，但要用不同的程序来进行算术运算。请试试用ⅩⅩⅥ（26）去除CⅣ（104）。

本书第一部分将考察在认知科学研究过去的40年中发展起来的研究心理表征和心理程序的不同途径。关于这些不同的途径的得失已有不少的争论，许多权威的认知科学理论家为他们所偏好的研究路线热烈地争辩。我个人采取的道路较为折衷，因为我认为目前已有的不同的心理表征理论其互补性大于其竞争性。人类的心智出奇的复杂，而我们对它的理解无论从它对规则的使用（如上述）还是从许多其他类型的（包括一些我们所不熟悉的）表征方式上都将获得收益。后者包括“联接主义者”（Connectionist）或“神经网络”式的表征，这些将在本书第七章予以讨论。

起　源

对心智及其活动进行理解的尝试至少可以追溯到古希腊，哲学家柏拉图和亚里士多德试图说明人类知识的本质。柏拉图认为最重要的知识来源于概念，比如美德就是这样的概念，人们不依赖于经验而能凭天赋获知。另外像笛卡尔和莱布尼茨这样的哲学家也相信只须借助于思维和推理就能获得知识，这种立场被称为**唯理性主义**（rationalism）。与之相反，亚里士多德通过“所有的人都是会死的”这样的规则来探讨知识。这种哲学立场，经由洛

克、休谟和其他一些哲学家的辩护，成为众所周知的**经验主义**(empiricism)。在 18 世纪，康德试图调和唯理性主义和经验主义，他主张人类的知识既依赖于感觉经验，也离不开心智的天赋能力。

直到 19 世纪实验心理学诞生以前，对心智的研究一直停留在哲学领域。威廉·冯特及其学生开创了较为系统地研究心理过程的实验方法。然而，在其后的几十年间，实验心理学逐渐为**行为主义**(behaviorism)所统治，行为主义实质上否定心智的存在。依照像 J.B. 华生 (Watson 1913) 这样的行为主义者的观点，心理学必须严格限于研究可观察的刺激与可观察的行为反应之间的关系，谈论意识和心理表征是为正当的科学研究所不耻的。尤其是在北美，行为主义统治心理学领域一直到本世纪 50 年代。

大约到了 1956 年，学术界的氛围开始发生剧变。乔治·米勒(Miller 1956) 总结了一系列研究，表明人类的思维能力是有限制的，例如对短时记忆来说，大约就限于 7 个条目 (这也就是为什么人们很难记住过长的电话号码和社会保障号码的原故)。他提出将信息编码分为组块 (chunk) 可以克服记忆容限，组块便是须要心理程序进行编码和解码的心理表征。与此同时，尽管第一代计算机才面世几年，约翰·麦卡锡、马文·明斯基、阿兰·纽威尔和赫伯特·西蒙便创立了人工智能这一研究领域。此外，诺姆·乔姆斯基 (Chomsky 1957, 1959) 强烈地抨击了行为主义关于语言是一种通过学习获得的习惯的主张，取而代之以由规则所构成的心理语法来说明人们理解语言的能力。上述六位思想家可以恰当地被称为认知科学的奠基人。

在后面的章节里我们将结合不同的心理表征理论勾画出认知科学在其后发展的历史。麦卡锡成为以形式逻辑为基础探索人工智能的领头人，我们将在第二章加以讨论。60 年代，纽威尔和西蒙揭示了依据规则来说明人类智能的威力，在第三章里我们要讨论延续这一传统所进行的一些工作。到了 70 年代，明斯基提出类似于概念的框架 (frame) 是知识表达的核心形式，而人工智能和

心理学领域的其他研究者也探讨所谓程式（schema）和脚本（script）这些与之相似的结构（第四章）。与此同时，心理学家们开始对心理表象显示出与日俱增的兴趣（第五章）。自80年代以来，更多的实验和计算研究集中于类比推理，也称作基于案例的推理（第六章）。80年代最激动人心的莫过于联接主义心理表征和加工理论的兴起，它大体上是以大脑的神经网络为模型的（第七章）。这些不同的研究途径都已对理解心智做出了应有的贡献，在第八章将对它们的长处和缺陷作出总结和评价。尽管如此，对于心智应当被理解为心理表征与程序这一核心论点仍存在着诸多挑战，这些将在本书的第二部分（第九章至第十一章）予以讨论。

认知科学的方法

认知科学绝不应当只是不同领域的人们在午餐会上聚在一起聊一聊心智是什么。但在我们着手审视认知科学中已取得的一致性的观念之前，我们先得鉴别一下不同领域的研究者对心智和智能的探索所带来的视角和方法的多样性。

虽然认知心理学家今天时常介入理论和计算模型的探讨，但他们首要的方法仍是以人为受试者进行实验。作为受试者的人，通常是符合实验要求的本科生，被带到实验室以便不同类型的思维过程能够在受控条件下得以研究。不妨举一些后面章节里将会提到的例子，心理学家通过实验检验人们在进行演绎推理时所犯错误的种类，人们形成和使用概念的方式，人们使用心理表象完成思考时的速度，以及人们运用类比求解问题时的表现。我们关于心智如何运作的结论绝不能仅仅建立在“常识”和内省的基础之上，因为它们提供的很可能是一幅引人误入歧途的图景，而且许多的心理过程是意识不到的。因此，从不同的方面慎重细致地揭示心理过程的心理学实验，是使得认知科学成为一门科学的关键所在。

虽然理论缺乏实验是空洞的，但没有理论的实验却是盲目的。为了弄清有关心智本质的根本性问题，心理学实验需要在一个以心理表征与程序为基础的理论框架中加以说明。而发展理论框架最好的办法之一是构造和测试计算模型，以便模拟心理运作。为了补充有关演绎推理、概念生成、心理表象和类比式问题求解的心理学实验，研究人员发展了各种计算模型来模拟人类完成这些活动的各个方面。设计、建造和测试这些计算模型是人工智能(AI)的核心方法，AI是研究智能系统的一个计算机科学分支。在认知科学中最理想的状况是，计算模型和心理学实验能够携手共进，但AI中很多重要的工作是在与实验心理学相对隔离的情况下检验不同的知识表达方式的能力的。

尽管有一些语言学家也做心理学实验或者研制计算模型，但大多数还是使用其他方法。对于乔姆斯基传统的语言学家来说，主要的理论任务是确定作为人类语言的基本结构的语法原则，这种鉴别任务要求对符合语法和不符合语法的言语之间的细微差别予以关注。例如，在英语中，句子“She hit the ball”和“What do you like?”是符合语法的，而“She the hit ball”和“What does you like?”则不是。英语语法就要解释为什么前者是可接受的而后者不是。在后面的章节中会给出乔姆斯基式的和其他学派的语言学家进行的理论和经验研究的例子。

同认知心理学家一样，神经科学家也进行受控实验，但他们进行的观察却大不一样，因为神经科学家直接关心大脑的本质。对非人类的受试者，研究者可以插入电极，记录单个神经元的激活。而对人类受试者，这项技术就太过于侵害性了。近年来采用磁扫描和正电子扫描装置，可以观察当人们在完成各种心理任务时大脑的不同部位所发生的情况。例如，脑扫描可以确定在运用心理表象或进行语词理解时大脑活动的区域。有关大脑功能的另一类证据来自于观察那些大脑受到某种方式的损伤后的人的行为。例如，发生中风时，大脑专管语言的一部分会导致诸如无法发出完

整句子这类的行为缺损。和认知心理学一样，神经科学既是理论性的又是实验性的，而理论的发展通常借助于研制反映神经元组群行为的计算模型。

认知人类学研究在不同的文化环境中人们如何思考，从而拓宽了对人类思维的审视视角。对心智的研究显然不应当只局限于讨论操英语的人们是如何进行思考的，而应当考虑到不同文化的思维模式可能的差别。在第十章将讨论认知科学如何越来越多地意识到需要在特定的社会环境和物质环境中考虑心智的活动。对认知人类学家来说，主要的方法是现场社群调查法，这就要求研究者同某一文化的成员共同生活，打成一片，这样才能把握他们所特有的社会和文化系统。例如，认知人类学家已经调查了不同文化中描述颜色的词语的相似和差异。

除了少数例外，哲学家一般不进行系统的经验观察或建造计算模型。但哲学对认知科学来说仍显得重要，这是因为认知科学涉及一些基本性的问题，这些问题对于心智的实验研究或计算研究的方式都是基础性的。像表征和计算的本质这种抽象问题在心理学和人工智能的日常研究中毋须顾及，但研究者对于他们的所做所为进行更深入的思考时就无法避免了。哲学还处理一些普遍性的问题，如心灵与肉体的关系，以及一些方法论问题，如认知科学中解释说明的实质。此外，哲学既涉及人们实际上是如何进行思维这种描述性问题，也关心人们应当如何进行思维这样一些规范性问题,这是因为除了理解人类思维这样的理论目标之外,认知科学还具有改善人类思维这样的实践目的，这就要求对我们想要的思维是怎样的进行规范性的反思。心智哲学本身并没有一种突出的方法，但应当与其他领域的最优秀的理论研究一同关注经验成果。

在其最弱的形式上，认知科学只是上述几个领域的汇合：心理学、人工智能、语言学、神经科学、人类学和哲学。当理论研究和实验研究在关于心智本质的结论上趋同聚合时，跨学科的研

究会变得更为引人入胜。后面的章节将会提供案例表明认知科学处于各个领域的交汇面上。从理论上看，目前最富有成效的方式是依据表征和计算来理解心智。

对心智的计算-表征理解

认知科学的中心假设是：对思维最恰当的理解是将其视为心智中的表征结构以及在这些结构上进行操作的计算程序。尽管在有关构成思维的表征和计算的实质是什么这一问题上存在争议，但这一中心假说本身足以涵盖目前的认知科学对思维的理解，包括联接主义理论。为简略起见，我将基于这一中心假说对心智的理解方式称作 CRUM（Computational-Representational Understanding of Mind）。

CRUM 可能是错误的。本书第二部分将展示对这一路线的一些根本性的挑战，对于解释有关心智的一些基本事实，表征和计算可能并不恰当。但通过评价各种知识表征理论所获得的成功，我们将会看到，在对心智的理解上CRUM 取得了可喜的进展。毫无疑问，CRUM 是至今为止在理论和实验上最为成功的方式。虽然在认知科学的诸领域中并非人人都赞同CRUM，但从心理学和其他领域中的权威性学术期刊上我们不难看出 CRUM 是目前认知科学的正统观点。

CRUM 的成功在很大程度上得益于从计算机的发展中获得了丰富的类比。在第五章中我们会谈到，类比对于提出新的科学思想颇有助益，而将心智比作计算机对于我们理解心智提供了一种强有力的方式，远远胜过了以往提出的诸如电话开关板之类的比喻。有计算机科学背景的读者对于计算机程序由数据结构和算法组成这样的概括一定很熟悉，现代编程语言都包含一系列的数据结构，包括像“abc”这样的字符串，像 3 这样的数字，以及更

为复杂的结构，如表（A B C）和树。算法——机械式的程序——可以定义为在各种数据结构之上的操作。例如，一段名为 REVERSE 的程序可以定义为转置一个表，将（A B C）变为（C B A）。这个程序可由两个更小的子程序组成，先从一个表中取出一个元素，再把它加在另一个表的开头，使计算机先得到（A），再得到（B A），然后是（C B A），从而得到一个倒置的表。同样，CRUM 假定心智具有心理表征，类似于数据结构，而计算程序则类似于算法。图解如下：

程　序	心　智
数据结构＋	心理表征＋
算法	计算程序
＝运行程序	＝思维

这就是认知科学中正统的类比，尽管它将另一个基于大脑的类比作了一点奇异的扭曲。联接主义者提出了关于表征和计算的新颖见解，将神经元及其联接比作数据结构，而将神经元的激活和激励传播比作算法。这样 CRUM 就是处于心智、大脑和计算机之间的一个复杂的三维类比。三者中的每一个都可以用以为其他两个提供新见解。并不存在某个独一无二的关于心智的计算模型，因为不同种类的计算机和编程方法决定了心智运作方式的不同。我们大多数人今天接触的计算机是串行处理器的，一次执行一条指令，但是大脑和某些最新开发的计算机是并行处理器的，可以一次做很多个操作。

如果你对计算机有相当的了解，从计算机来理解心智就来得相当自然，即便你不一定同意心智从根本上说是一台计算机。从未编写过计算机程序但使用过食谱的读者可以考虑另一个类比。一个食谱通常有两个部分：一张调料表和一套调配调料的指令集。一盘菜是将烹调指令运用于调料的结果，恰如一段可执行的程序

是将算法运用于诸如数字和表这样的数据结构的结果。将菜谱来比思维有些牵强，因为调料不是表征，而烹调指令需要有人对其进行解释。第二章至第七章会提供一些更为直接对应于心智操作的计算程序的例子。

理论、模型和程序

对于心理过程的理论研究来说，计算机模型通常是十分有用的。对认知科学模型的把握要注意四个方面的区别与联系：理论、模型、程序和平台。一个认知**理论**要假定一套表征结构和一套在这些结构上进行操作的加工过程。通过与计算机程序由数据结构和算法构成进行类比说明，一个计算**模型**使得这些表征结构和过程更为精确。有关表征的模糊概念由准确的关于数据结构的计算概念加以补充，而心理过程则可由算法来定义。为了测试该模型，必须用一种编程语言（比如 LISP 或 C）将其在一个软件**程序**中实现。此程序可以在各种硬件**平台**上运行，例如 Macintosh、SUN 工作站或 IBM PC，或是某种专门设计的专用硬件设备，如一台带有多个并行处理器的 Connection Machine。各种结构和加工过程均可依此方法进行研究，从一些传统人工智能的规则和搜索策略，到较新的联接主义的分布式表征和传播激励加工。

举例来说，例如要了解儿童怎样学习把两个数相加，如 13+28 这样的问题。一个认知理论就需要指出儿童如何表征这些数字以及怎样加工这些表征以完成加法运算。这个理论要指出 13 是由单一结构来表示，还是由诸如 10 加上 3 这种复合结构，或是一种类似神经元的复杂结构来表示的。该理论还要指出在这些结构上操作以得到结果 41 是怎样一些过程，包括设法将 30 加上 11 转变成为 41 的移位操作。一个计算模型通过刻画可编程的结构和算法从而更为详尽、准确地说明表征和加工处理的实质，使之与做加法的心理表征和加工过程相类似。为了评估该理论和模型，我们

要用某种计算机语言如 LISP 来编写一个计算机程序，并运行这个程序以便同人做加法的行为进行比较，不仅要看这个程序能否跟人一样得出正确的答案，还要看它是否会犯与人同样的错误。这个程序也许能在多种不同的平台如 PC 机上运行，或者是为某种特殊的计算机（如一台模仿大脑神经元结构的计算机）而专门编制的。

在认知理论发展的三个阶段：发现、修正和评估，心智与计算机的类比都能发挥作用。与不同类型的程序相关的计算概念通常能为提出新的心理结构和加工过程提供启示。理论、模型和程序的发展通常是携手并进的，因为编写程序可以导致发明新的数据结构和算法，而后者又可能成为模型的一部分并在理论中找到对应物。举例来说，在编写计算机程序模拟人做加法时，编程者可能会想到一种新的数据结构而对儿童如何表征数字提供启发。同样，在对理论、模型和程序进行评价时，三者通常是缺一不可的，因为我们对理论的确信依赖于通过程序的性能所反应出的模型的有效性。如果说一个做加法的计算机程序无法完成加法运算，或者它做的加法比人类更为完美无缺，我们就有理由相信与该加法程序相应的认知理论是不适当的。

对于评价模型和理论，运行的程序可以在三个方面做出贡献。首先，它有助于显示提出的表征和加工过程是否是计算可实现的。这一点很重要，因为很多算法初看起来显得很合理，但在计算机上却无法扩展到较大的问题上。其次，一个理论不仅要有计算上的可实现性，还必须具备心理学上的合理性。通过将程序应用到各种思维的实例上，可以得到定性的反映。例如，加法的程序应该得到同儿童做加法运算同样的正确和错误的答案。第三，为了反映理论与人的思维能否在更多的细节上相吻合，程序可以对人的思维作出详细的定量化的预测，与心理学实验的结果进行对比。如果说心理学实验表明儿童在完成某类相加运算时正确率有一定的百分比，那么计算机程序应该得到大致相同的正确率。认知理

论本身一般不足以精确到产生出这样的定量化预见，但模型和程序可以填补理论与观察之间的距离。

对心理表征的评价方法

现在我们要谈一谈对于一个心理表征理论的要求是什么。框盒1.1列举了对于说明思维的某一特定的表征与计算理论进行评价的五个复杂的标准。在第二章至第八章我们将采用这些标准来评价六种说明心理表征的方式：逻辑、规则、概念、表象、案例和联接（人工神经网络）。

框盒1.1 评价心理表征理论的标准

1. 表征力
2. 计算力
   a. 问题求解
      i. 规划
      ii. 决策
      iii. 解释
   b. 学习
   c. 语言
3. 心理学上的合理性
4. 神经学上的合理性
5. 实践上的可应用性
   a. 教育
   b. 设计
   c. 智能系统

在第二章至第七章中介绍的六种方式都提出了一种特定的表

征以及一套相应的计算程序。第一条标准，表征力，涉及一种特定的表征方式能表达多少信息。例如，校历上强调："一旦被录取，学生应在课程开始之前预先登记欲修的课程。"重视这条通知的学生就需要以一种内在形式来表达这一通知，以期作出进一步的推论，比如得出他们应到注册办公室去登记新学期的课程这一结论。我们将会看到不同类型的心理表征方式在表征力上是大不一样的。

心理表征之所以重要，不仅在于它们能表达什么样的事物，对于你能用它们来做什么事也是十分关键的。我们可以从如何说明三种重要的高层思维活动来评价某种心理表征方式的计算力。首先是问题求解：一个心理表征理论应当能够说明人们怎样推理以达到其目的。至少有三种类型的问题求解需要得以说明：规划、决策和解释。规划要求推理者能够解决怎样从一个初始状态经由一系列中间状态而到达目标状态。规划的问题包括从怎样在你的航班起飞前到达机场这样的现实问题，到学生在课本上和考试中遇到的练习题。对这些问题，先会给定一些信息，然后要求同学们求得正确答案。初始状态包括学生所知道的背景知识，以及问题描述中的信息，而目标状态则包括正确答案。学生要通过一系列成功的计算步骤来找到答案。

在决策制定中，人们面临的问题是有多种途径来完成任务，而此时要求选择一种最佳途径。例如，一个即将毕业的学生面临的选择有：找工作、升人研究生院或者去上法律或商业等方面的职业学校。这样的选择是非常困难的，因为这需要学生能确立自己的目标并找到实现这些目标的最佳途径。规划问题的任务是找到一系列成功的行动，而决策问题则是在多种可行的方案中选择最佳方案。

解释问题要求人们对某些事情为什么会发生给出说明。这些问题的范围从一个朋友为什么昨天晚餐迟到这样的日常生活问题，到为什么人类的语言会演变这样深奥的科学问题。每一个有

起码的聪明才智的人都有能力进行规划、决策和解释。一个认知理论必须有足够的计算力对于人们如何解决这些类型的问题给出可能的说明。

一个表征和加工系统的计算力不仅在于该系统能够完成多少计算量，同时还要考虑到其计算的有效性。假定有一个程序运行一次需用 2 秒钟，运行第二次所用的时间加倍，运行第三次时间再加倍，如此下去，运行 60 次则需 2^{60}秒，这比宇宙形成 200 亿年的时间还要多了。不管是天然的还是人工的智能系统都必须有足够的速度，这样在它们所处的环境中才能有效的工作。

人们在解决问题的时候，通常能够从经验中学习，从而在下一次的时候能够更容易地解决问题。例如，学生在第一次选课时经常由于对所应遵循的程序或者对如何选择较好的课程都不甚了解而感到迷惑，但后来选课登记就变得容易多了。具有智能就包括能从经验中进行学习，所以一个关于心理表征的理论必须具备足够的计算力来解释人们是如何学习的。在后面讨论心理表征的不同方式时，我们会看到有多种不同的人类学习方法，从获得如“注册”这样的新概念和“千万别选上午 8 点 30 分的课”这样的规则、到更微妙的对行为的调整。

除了问题求解和学习，一种普遍性的认知理论还必须对人们的语言运用给予说明。人类是地球上唯一能使用复杂语言的物种，问题求解和学习的一般原则可能可以说明语言的使用，但也有可能语言是一种须特别对待的独一无二的认知能力。语言的使用至少有三个方面须加以解释：人们理解语言的能力、人们表达发声的能力以及儿童学习语言的普遍能力。不同的知识表征方式对这些问题给出了不同的回答。

如果将人工智能视为一门工程学科，它完全可以在忽视人类是如何完成各种任务的前提下去发展有关问题求解、学习和语言的计算模型，因为对人工智能来说，问题是怎样使得计算机能做这些事情。但认知科学的目标之一是理解人类的认知，所以它要

求一个心理表征的理论不仅要有足够的表征力和计算力，还必须关注人们是怎样进行思维的。这样，评价一个心理表征理论的第三个标准就是心理学上的合理性，这就要求不仅从定性的角度要说明人类的各种能力，还须考虑涉及这些能力的心理学实验的定量结果。这些实验包括上面在讨论计算力时所提到的那些高层任务：问题求解、学习和语言。心理学标准和计算力标准的差别不仅在于心理表征的认知理论必须在计算上是可能的，还要能够说明人类在完成这些任务所采用的特定方式。

同样，由于人的思维是由人的大脑来完成的，心理表征的理论至少要与神经科学实验的结果相一致。如果说以往的一些神经科学上的技术如脑电波的 EEG 记录对于我们了解高层认知活动还过于粗略的话，那么近十年来发展起来的新的扫描技术可以使我们确定在完成某些特定认知任务时大脑活动的位置和时间。认知神经科学由此而成为揭示心智活动的一个极其重要的部分，因此我们应力求从神经科学上的合理性来评价每一种知识表征方式，尽管目前关于大脑如何进行认知的信息仍相当有限。

评价心理表征理论的第五个标准即最后一个标准是实践上的可应用性。虽说认知科学的主要目标是理解心智，但这一理解可以导向许多可望的实践成果。本书对六种知识表征方式考察三种重要的应用：教育、设计和智能系统。对教育而言，认知科学可以增进我们对于学生是如何进行学习的了解，这对改进教学会有所启发。像怎样使得计算机界面更为人们乐于使用这类设计问题，也可以从对人们在执行任务时如何进行思考的理解中获得助益。最后，不管是建造独立的专家系统还是作为人类决策的支持工具，智能系统的开发都可以从对人们怎样思考的研究中得到帮助。不同的心理表征理论可以指导开发出不同类型的专家系统，包括基于规则的、基于案例的和联接主义的工具。认知科学的其他可能的应用还包括对心理疾病的研究与诊治。

我们将会看到，没有一种心理表征方法能完全满足上述的所

有标准。此外，还有人类思维的一些其他方面，如知觉（视觉、听觉、触觉、嗅觉、味觉）、情绪和运动控制，未被包含在这些标准中（见第九章）。然而，这些标准毕竟为比较与评价目前的各种心理表征理论提供了一个框架，既考虑到了它们所取得的成就，也没有回避它们的不足之处。

小　　结

心理学、人工智能、神经科学、语言学、人类学和哲学的研究者在探索心智的过程中所采用的方法各不相同，但这些不同方法在对心智是如何工作的解释上有一些共同之处。一个对认知科学的统一的视角是将各种理论途径都视为对心理表征和心理程序的关注，这与人们熟悉的计算机编程中的表征和程序是相似的。对心智的计算-表征理解可由下面的解释程式来概括：

解释目标

为什么人们会有某种特定的**智能行为**？

解释模式

人们具有心理**表征**。

人们具有在这些**表征**上进行操作的算法**程序**。

这些**程序**，运用到**表征**上，产生出**行为**。

用黑体字印刷的词是可代入置换的，针对解释不同类型的智能行为，可以代入不同类型的表征和程序。当前，有六种主要的方式来建造心智的模型，包括逻辑、规则、概念、类比、表象和神经联接。可以根据五个标准对它们进行评价：表征力、计算力、心理学上的合理性、神经学上的合理性和实践上的可应用性。

下面是指导本书撰写的一些基本预设：

1. 对心智的探索是激动人心而且至关重要的。这一研究在理论探索上令人兴奋，是因为对心智本质的探索与科学上对其他事物的探索一样，都是富有挑战性的。这一研究在实践上亦有其重要性，因为了解心智是如何工作的对许多不同的领域都是有必要的，诸如改善教育、改进计算机和其他产品的设计，以及开发专家系统等。

2. 对心智的探索是跨学科性的。它需要来自哲学家、心理学家、计算机科学家、语言学家、神经科学家、人类学家以及其他方面的见解。不仅如此，它还需要这些领域所发展的多种多样的方法学。

3. 对心智的跨学科研究（即认知科学）有一个内核：心智的计算-表征理解(CRUM)。思维是心理表征以及在这些表征上的计算过程的操作结果。

4. CRUM 是多种多样的。对于理解人类思维有多种表征和计算，目前还没有一种单一的计算-表征能涵盖人类思维的整个领域。本书（第二章至第七章）介绍了根据表征和计算来理解心智的六种主要的途径。

5. CRUM 是成功的。在说明心理学性能的理论能力方面和提高这些性能的实践能力方面，计算-表征方式都超过了它以前的所有关于心智的理论。

6. CRUM 是不完善的。并非人类思维和智能的所有方面都能完全由计算-表征来加以说明。对 CRUM 的实质性挑战表明了将它与生物学方面的研究（神经科学）以及思维和知识的社会性方面的研究进行整合的必要性。

讨论题

1. 举例说明，在学生进入大学后，还有哪些东西需要学习？

2. 为什么不同领域的研究人员在研究心智时采用不同的方

法？

3. 对于理解心智，你能想出不同于计算-表征的方式吗？

4. 人类思维的哪些方面是计算机最难以执行或模拟的？

5. 认知科学中的理论和模型与物理学或其他领域中的理论和模型相似吗？

6. 你认为一个心理表征的理论是否还需要满足其他的标准？

进一步的推荐读物

关于认知科学的历史，参见 Gardner 1985 和 Thagard 1992，第九章。其他的认知科学导论书籍包括 Johnson-Laird 1988 和 Stillings 等人 1995。综合性的论文集有 Osherson 1995 和 Posner 1989，

认知心理学的课本有 Anderson 1990 和 Medin 与 Ross 1992。人工智能的导论书籍，见 Rich 与 Knight 1991 和 Winston 1993。心智哲学的导论有 Bechtel 1988 和 Graham 1993。Akmajian 等人 1995 是一本导论性的语言学教材。Kosslyn 和 Koenig 1992 提供了一本可读性强的认知神经科学的入门书，而 Churchland 和 Sejnowski 1992 的则更偏重于计算方面。认知人类学的导论书籍有 D'Andrade 1995。

备　　注

对思维即计算的讨论通常会提到一种抽象的计算模型，比如图灵机(Turing Machine)，这是一种由纸带和一个可在纸带的空白处书写符号的读写头所组成的简单装置。虽然可以从数学上证明这样一种机器原则上可以做其他任何计算机能够做的任何事，但作为与思维的类比，图灵机就显得过于抽象了；而从高层的计算概念如数据结构和算法来展开讨论就好得多。有关图灵机的描述，见第十章。

有关解释程式（schemas）和模式（patterns）的更多的材料，见 Kitcher 1981，1993，Leake 1992 和 Schank 1986。

第二章 逻　　辑

尽管形式化的逻辑并非通向心理表征最具影响力的心理学途径，但我们仍有足够的理由从它展开我们对心理表征的探讨。首先，有关表征和计算的许多基本概念来自逻辑传统。其次，当前许多哲学家和人工智能研究者将逻辑视为推理的核心。第三，逻辑所具有的充实的表征能力足以令其与其他具有更强的计算有效性和心理学合理性的心理表征方式一争高下。

形式逻辑肇始于2000多年前的古希腊哲学家亚里士多德。他系统地研究了下述的推理形式：

所有的学生都负担过重。
玛丽是一名学生。
所以，玛丽负担过重。

这样的推理模式，由两个前提和一个结论组成，称为**三段论**。通过对许多不同类型的三段论进行分类整理，亚里士多德揭示了怎样能纯粹地根据它们的形式进行分析。对上面例子中由两个前提所推导出的结论而言，这个三段论的内容是否关于负担过重的学生是无关紧要的。我们可以用“香肠”来替换“学生”，用“橙子”来替换“负担过重”，用“马文”来替换“玛丽”，这样我们可以从修正过的前提推出马文是橙子这一结论，尽管这个结论没什么意思。亚里士多德引人了符号的使用来展示推理的形式：

所有的 S 是 O。
M 是 S。

所以，M 是 O。

亚里士多德作出的这一发现，即怎样完全从形式上分析三段论，而不计其内容，对后来的逻辑研究产生了深远的影响。不过，从心理学的角度看，这一发现的效用受到了挑战，在下面有关心理学上的合理性一节里我们将会看到。

三段论是**演绎**推理的一种形式，其结论必然地从前提中推导出来：如果前提为真，则结论也为真。**归纳**推理由于引入了不确定性就危险多了。假如你所认识的所有学生都负担过重，你可能归纳地推出所有的学生都负担过重。但你的结论很可能是错的，比如说如果有篮子编织专业的学生，你可能就不知道他们学得很轻松。

尽管三段论在2000多年的时间里统治了对逻辑的研究，它却不足以表达所有的推理方式。对一些简单谓词如“是一名学生”，三段论可谓胜任愉快，但它却无法处理像语句“学生选课以获得学分”中的“选”这样的关系。在此选是学生与课程之间的关系。1879年，奥地利数学家高特罗伯·弗雷格（Frege 1960）开创了现代逻辑，他设计了一个比亚里士多德系统更为通用的形式化逻辑系统。随后，伯特兰·罗素和许多其他逻辑学家找出了很多办法来增强形式逻辑的表征与演绎能力。

逻辑学家阿隆佐·邱奇和阿兰·图灵创立了早期的计算理论。30年代，邱奇、图灵和其他学者发展了能够说明什么是有效计算的数学模型。这些模型在数学上是相互等价的，这对将我们关于有效可计算性的直觉观念与清晰定义的数学概念（如图灵机可计算性）等同起来提供了支持。当40年代末50年代初数字计算机面世时，关于可计算性的数学理论为人们理解计算机的运作提供了一种有力的工具。因而，就不难理解在50年代中期人工智能诞生时，像约翰·麦卡锡这种数学出身的学者会将逻辑视为最恰当的工具。不过，我们也将会看到，人工智能的其他先驱者：阿

兰·纽威尔、赫伯特·西蒙和马文·明斯基所偏好的是其他的方式。

表征力

现代形式逻辑具备很多种演绎推理的能力。最简单的形式逻辑是命题逻辑，在命题逻辑系统中可用表达式 p 和 q 来代表像“波拉在图书馆”和“昆西在图书馆”这样的语句。简单表达式可以使用符号组合成较为复杂的表达式，如用“&”代表“与”，“∨”代表“或”，“→”代表“如果-那么”。例如，语句

如果波拉在图书馆，那么昆西也在图书馆。

就变成了

$p \rightarrow q$。

这样的“如果-那么”称为条件句。为了表达否定，“非 p”可以写成“$\sim p$”。用这些建筑材料我们便可以建构起复杂陈述句的形式化表达。像“如果波拉在图书馆或昆西在图书馆，那么德博拉就不在图书馆”就可以形式化为

$(p \vee q) \rightarrow \sim d$。

在此，“p”代表“波拉在图书馆”，“q”代表“昆西在图书馆”，而“d”代表“德博拉在图书馆”。

复杂一些的逻辑可以发展成不同类型的命题算子。模态逻辑增加了表示必然性和可能性的算子，这样我们可以表示“波拉可能在图书馆”这样的陈述句。认识逻辑加入了代表知识和信念的

算子，这样可用“Kp”表示“已知 p”。道义逻辑可以表达诸如 p 是准许的或是被禁止的这样的道德观念。

命题逻辑要求将诸如“波拉是一名学生”这样的陈述句视为一个不可分割的整体，但谓词逻辑则允许我们将它拆开。谓词演算区分谓词和常项，前者如“是一名学生”而后者则是指波拉和昆西这样的个体。通常在哲学课程里的谓词逻辑中，“波拉是一名学生”被形式化为“$S(p)$”，这里“p”代表波拉而不是一个完整的命题。计算机科学家则倾向于表达为更有助于记忆的形式，如“是学生（波拉）”。除了简单的属性外，谓词还可以用于表示两件或更多事物之间的关系。例如，“波拉选修 PHIL256”变为“选修（波拉，PHIL256）”。

通过引入变量如“x”和“y”，谓词演算可以用量词如“所有的”和“一些”对句子进行形式化。例如，“所有的学生都负担过重”就变成了：

（对所有的 x）（是学生（x）→负担过重（x））。

在字面上，这是说“对任何 x，如果 x 是一名学生，那么 x 负担过重”，这等于说所有的学生都负担过重。句子“选课的学生会得到学分”可以形式化为：

（对所有的 x）（对所有的 y）[（是学生（x）& 课程（y）& 选课（x,y））→得到课程学分（x,y）]。

乍看起来挺复杂的，说起来就是“对任何的 x 和 y，如果 x 是一名学生，y 是一门课程，且 x 选修 y，那么 x 就会从 y 获得学分。”

现在，兴趣偏重于心理学的读者也许会问了：你为什么把这么一大堆数学符号扔给我？答案是：形式逻辑的一些基本知识对于理解当前认知科学的许多工作是必需的，比如一些关于人们如

何进行演绎推理的理论。至少，我们应注意到人们能理解像“通过了课程要求的学生可由此得到学分”这样的陈述并以此来作推理。谓词逻辑，不同于我们将要讨论到的一些其他的表征方式，具有足够的表征力来处理这样的情况。

虽说谓词逻辑对解决许多问题来说都有用途，但它也有其局限性，特别是我们要将一个自然语言的文本翻译为谓词逻辑系统时就显得特别突出。例如，试把上面的一个自然段变成逻辑形式。第一个句子里有“现在”一词，而谓词逻辑要扩展到能处理时间，就不是一件容易的事。第一句话里还有“你”这个词，读者很清楚这是指保罗·萨伽德，本书的作者，但如何用逻辑来表示这一点就不太清楚了。此外，第一句的结构中包含了关系“问”，这涉及一个提问者和他（她）所问的命题，这样我们就必须能在一个命题中嵌入另一个命题，这用一般的谓词逻辑的形式化方法来做就很不自然了。假如从语言到逻辑形式的转化能容易一些的话，也许我们会有信心认为形式逻辑具有对心理表征来说所必要的一切优点。

命题逻辑和谓词逻辑在做断言时，即陈述句要么为真要么为假，显得得心应手，但若引入了不确定性，比如“波拉可能在图书馆”，它们就无能为力了。对这样的断言，可以用概率论来补充形式逻辑，将 0 到 1 之间的数字分配给命题，这样我们可用“$P(p)=0.7$”来表示波拉在图书馆的概率是 0.7。

计 算 力

表征本身并不能做任何事情。为了能够进行思维，必须在表征之上能够进行操作。为了推导出逻辑上的结论，我们须将推理的规则运用到一个前提的集合上去。两种最常见的推理规则能使我们从条件句（如果-那么语句）推出结论。

肯定式（Modus ponens）

$p \rightarrow q$

p

所以，q 。

否定式（Modus tollens）

$p \rightarrow q$

非 q

所以，非 p 。

从条件“如果波拉在图书馆，那么昆西在图书馆”，以及波拉在图书馆的信息，肯定式使你能推出昆西在图书馆。从昆西不在图书馆的信息，否定式可得出波拉也不在图书馆的结论。

在谓词逻辑里，有推理规则用于处理量词“所有的”和“一些”。例如，全称例示（universal instantiation）规则可以使我们从一个普遍陈述推导出一个例证，从（对所有的 x ）（冷漠（ x ））到（冷漠（波拉））是成立的，即从“每件事物都是冷漠的”到“波拉是冷漠的”。稍复杂一点的应用可用于“所有的学生都负担过重”这样的概括语句上：（对所有的 x ）（是学生（ x ）→负担过重（ x ））。将它用到玛丽身上，我们得到的结论是：如果玛丽是一名学生，那么她的负担过重：是学生（玛丽）→负担过重（玛丽）。

抽象的规则如肯定式本身并不能处理操作。为了能完成计算，它们必须成为人或者机器系统的一个部分，该系统能够将它们以适当的逻辑形式运用到语句上。从逻辑的角度看，演绎推理使用形式化的推理规则，只考虑前提的逻辑形式。

问题求解

规划 很多的规划问题可以由逻辑演绎来解决。假定泰芬妮是一名学生，她想获得心理学的学位。从她所在大学的手册上她知道她必须选修10门心理学课程，其中包括两门统计学课程STAT1和STAT2。在选修STAT2之前必须先修STAT1，而STAT2又是在一门必选的研究方法课程之前必须预修的。从手册上的普遍性描述，泰芬妮可以用全称例示规则将这些条件用到她本人身上：

选修（泰芬妮，STAT1）→可选（泰芬妮，STAT2）

可选（泰芬妮，STAT2）&开课（STAT2）→选修（泰芬妮，STAT2）

选修（泰芬妮，STAT2）→可选（泰芬妮，研究方法）

可选（泰芬妮，研究方法）&开课（研究方法）→选修（泰芬妮，研究方法）

选修（泰芬妮，研究方法）&选修（泰芬妮，STAT1）&选修（泰芬妮，STAT2）&选修（泰芬妮，其他7门课）→毕业（泰芬妮，心理学学位）。

最后一个条件句是对下面陈述句的形式化：如果泰芬妮选修了研究方法课，两门统计学课和七门其他课程，那么她就能获得心理学学位毕业。泰芬妮便可以使用这些条件句和肯定式推理规则来制定一个规划，用逻辑术语来说就是从她的初始状态，即她未修过任何一门心理学课程，到目标状态，即她毕业，所进行的一个演绎推理。泰芬妮便可以构造这样一个演绎规划：她先选修STAT1，再修STAT2，然后是研究方法课，再后是七门其他课程，最后获得心理学学位从而毕业。

为了使得规划能够在计算上可实现，演绎必须比形式逻辑里的一般性推理规则增加更多的约束。例如，命题逻辑里含有下面的合取规则：

合取

p

q

所以，$p \& q$ 。

此规则在逻辑上是优雅的，但在计算上却具有潜在的灾难性。如果泰芬妮已选修了两门统计学课，她可以这样推断：

选修（泰芬妮，STAT1）& 选修（泰芬妮，STAT2）

但若她增加下面的有效推断却不会增加任何新东西：

选修（泰芬妮，STAT1）& 选修（泰芬妮，STAT2）& 选修（泰芬妮，STAT1）& 选修（泰芬妮，STAT2）

这种未经控制的推理会很快耗尽任何人或机器系统的记忆。

规划的演绎方法在直觉上很吸引人，但它会遇上一系列计算上的问题。首先，它导致速度上的低效，虽说已经发展了很多计算策略以使演绎更有效，但哪怕一个简单的规划也需做大量的推理。其次，纯粹的演绎规则是**单调的**：它只能导出新的结论而不能否定先前的结论。（一个单调的数学函数是指函数值连续地增加或减少而没有波动；日常的推理不是单调的，因为有时旧的信念必须抛弃掉而不会持续地增加新的信念。）人工智能的研究者们已经发展了好些技术使得逻辑成为非单调的，但这在计算上的开销甚大。第三，一个完全的演绎型规划者不能从经验中学习。在解

决了一个问题之后，再次遇到这样的问题还会再从头来一次繁琐的演绎过程，除非是加进了一些从经验中学习的方法。

上面我仅是浮光掠影地介绍了人工智能关于演绎规划的工作（详细的介绍见 Dean 和 Wellman 1991）。读者应能看到逻辑演绎是描述规划问题如何解决的一种有用的方式，但以此来进行规划则存在一些困难。后面的章节里会介绍进行规划的其他途径。

决策 演绎式规划寻找的是一条从初始状态到目标状态的逻辑途径。然而，当出现合理途径不只一条的情况怎么办呢？在上一节的例子里，泰芬妮的演绎规划是先修 STAT1，然后是 STAT2，再学研究方法。但她经常会遇到在不同的行动中面临抉择的情况。例如，她可能会被要求选修一门人文学科的课程，这样她就必须在哲学、英语和西班牙语中进行选择。演绎式规划并不能告诉她选择哪一门，因为选择哪一门都能够到达满足人文学科课程要求的目标状态。泰芬妮需要判断哪一门课能够满足她的其他目标，比如学些有趣味的东西，不太难，以及所选的课程能够适合她课程表上的其他安排。演绎式推理可以有助于得出某些可能的选择方案。如果西班牙语只在上午 8：30 开课，泰芬妮就会演绎地推断出如果她选了这门课就必须早起。但其他一些结论可能就不那么清楚了，因为学生们通常不太了解某门课究竟怎么样。

因此决策的制定经常要考虑到概率。泰芬妮也许会认为哲学可能比英语有趣味，或者相反，或者西班牙语可能比英语更有用一些。因此她必须在确定了她的目标是什么以及对所选择的行动能达到目标的概率进行了估计的基础上作出决策。我们用“$P(p/q)$”来表示在条件 q 下 p 发生的概率，这样“P（有趣味的课程/英语课）”可以表示泰芬妮选修英语课而学到一门有趣味课程的概率。为了估算这一概率，她可以使用到在她的学校里有趣味的英语课所占的比例这样一些背景知识。为了决定究竟是选

修哲学、英语还是西班牙语，泰芬妮要计算出每一种选择的期望值，要考虑到各种结果出现的概率，以及她的目标在何种程度上得到满足。

已经有人开发出了基于概率的计算系统。赫兹曼（Holtzman 1989）运用概率论和其他形式化概念研制了一个智能决策系统，以帮助不育夫妇决定选用哪种治疗方案。开发概率计算机系统很需要技巧，因为使用概率可能会导致计算爆炸：随着模型中命题或变量数目的增长，所需概率的数目可能会呈指数增长。采用一些巧妙的技术可使得概率推理可计算（Neapolitain 1990；Pearl 1988）。下面会解决的另一个问题则是人们正常的决策制定过程中是否使用了概率。

解释　在规划问题中人们试图去解决怎样实现一个目标，而解释则是试图理解为什么某些事情会发生。假定萨拉与弗兰克约好在学生酒吧见面，而他却没有来。萨拉自然会对弗兰克的缺席设想一个解释，与规划相似，有时候解释可以看成逻辑演绎：你会从你所知道的去推断你想要解释的。有人告诉萨拉，弗兰克正为一门考试做准备，而他专心学习时他会忘掉约会。从这条信息萨拉便能解释为什么弗兰克会爽约了。

这一将解释视为逻辑演绎的观点由科学哲学家卡尔·亨佩尔提出并为之进行辩护（Hempel 1965）。特别是在像物理学这样的数理科学领域，解释可以说成是逻辑演绎。不过，在后面的章节里，我们会看到并非所有的解释都是演绎式的。例如，根据三角学和光学原理，从旗杆阴影的长度我们可以推算出旗杆的高度来，但如果说是旗杆阴影的长度解释了旗杆的高度，这就显得离奇了。

在少数情况下，弗兰克爽约的原因可以推演出来。举例来说，如果弗兰克是一个极严格的人，只有当他生病的时候他才可能不赴约。这时萨拉便可使用肯定式：如果弗兰克爽约了，他必是生病了；弗兰克这次未能赴约；所以弗兰克一定是生病了。但通常

情况下会有不止一个可行的解释，这就像一个规划者可以找到好几条通往目标的途径。萨拉也许可以构想出好几种演绎式解释，看下面的条件句：

如果弗兰克病了，那么他不会来赴约。
如果弗兰克出车祸了，那么他不会来赴约。
如果弗兰克爱上别人了，那么他不会来赴约。

如果萨拉实际上并不清楚弗兰克是否病了，或者出车祸了，或者爱上别人了，那她不会立即推断出他不会来赴约。但这三个条件句可用于对所发生的事构造假说：也许是他病了，也许他发生了车祸，也许是他爱上了别人。这种推理，你提出一个假设以便得出一个解释，被 19 世纪的美国哲学家查理士·皮尔斯（Peirce 1992）称为**逆推**（abduction）。萨拉可以根据这个假设反推出弗兰克病了，加上她知道如果弗兰克病了就不会赴约这一规则，这样她便可以演绎地解释为什么弗兰克没有来。逆推推理不能保证结论正确，因而是有风险的，但不失为一种有力的学习方式。

学　　习

智能系统不仅要能够解决各种问题，还应当能利用经验以改进其性能。我们怎样才能增强规划、决策和解释的能力呢？在逻辑的范围内直接改进问题求解能力所作的工作还不多，但逻辑表征对描述一些学习程序还是十分有用的。

不妨看一下在学生刚入校园时所面临的学习问题。学生们往往对所开设的课程以及在校园里会结识的人知之甚少，但很快通过接触各种具体的课程和各种类型的人，有关信息会不断增加，而且会很自然地对这些实例进行归纳概括。这些未经加工的概括可能包括哲学课很有趣（或很乏味），以及统计学课是必修课等。这些概括是归纳性的，因为在其中引入了不确定性，这是从确切知

道的知识到最有可能的结论之间的一个跳跃。学生在选修过两门哲学课程后，可能会作出下面的概括，用逻辑形式表达就是：

有趣（PHIL100）
有趣（PHIL200）
因此，（对所有的 x）（是哲学课（x）→有趣（x））。

其结论是所有的哲学课都有趣。但很可能这两门哲学课有趣而其它的哲学课（比如篮子编织的哲学）就很乏味。

归纳概括的计算机程序一般不用逻辑表达作为输入。昆兰(Quinlan 1983）的 ID3 程序是被使用得最多的学习程序之一。由于他在从示例集合形成概括时使用了概率,可以划归逻辑方式。例如，给定一组不同专业学生的样本以及对他们特点的描述，它可以概括出有关不同领域（如文科、理科和工科）的学生在个人特点、社会关系和智力上的不同点。与归纳概括相似而不同于演绎推理，逆推也是一种明显具有风险性的推理形式。萨拉在解释为什么弗兰克在与她约会缺席时，很可能还有许多她所不知道的解释。但逆推在科学上及人们的日常生活中却是不可或缺的，不论是古生物学家试图解释恐龙为什么会灭绝还是学生试着理解他们朋友的行为。由于逆推的目的是作出解释，而解释有时可以根据逻辑演绎来理解，所以将逆推置于一个逻辑框架内就很自然了（参见 Konolige 1992）。后面的章节里会介绍其他看待逆推的方式。

萨拉并不是想得出关于弗兰克爽约的一些解释，她要的是一个最好的解释。从逻辑的角度看，评估最佳解释要引入概率。针对弗兰克未赴约，萨拉要能估计出弗兰克生病的条件概率，以及所有其他假设的条件概率。概率计算的一个定理——贝耶斯定理——非常有用。简言之，贝耶斯定理指出，在给定条件下一个假设发生的概率，等于该条件的先验概率 $P(h)$，乘上该假设下的条

件发生的概率，再除以该条件本身的概率。以概率方式来解决怎样选择解释性假说在人工智能（见 Pearl 1988）和哲学（Howson 和 Urbach 1989）上是很普遍的，但其他方式也是存在的，我们将在第七章讨论。

“归纳”一词的使用非常混乱，因为它有广义和狭义两种含义。广义归纳覆盖了除演绎推理以外，任何引入了不确定性的推理。狭义则仅指归纳概括，从个别事件中得出普遍性结论。逆推（形成解释性假说）是广义的归纳而非狭义的。在本书中，我的习惯是使用“学习”来指广义归纳而用“归纳概括”来指狭义归纳。有关学习的其他计算上的说明会在其他章节里碰到。

语　言

语言学家有时会将形式逻辑视为理解语言结构的天然工具。甚至有一册再版书，书名就叫《语言学家想要知道的逻辑大全——而他们却耻于发问》（McCawley 1993）。哲学家理查德·蒙太古（Montague 1974）主张自然语言和人工语言在理论上没有什么重要的区别。但大多数语言学家和心理学家都不赞同这一主张，认为形式逻辑在理解人类语言时仅扮演一个小角色而已。斯塔伯勒（Stabler 1992）用逻辑对乔姆斯基的一些关于语言的最新见解进行了形式化处理，乔姆斯基提出了一个“逻辑形式”的层面，在这个层面上意义（meaning）能得到最清晰的表达（Chomsky 1980）。后面的章节会谈到其他类型的表征方式，尤其是规则和概念，怎样用于描述和解释人类对语言的使用。

心理学上的合理性

历史上，逻辑学家对于逻辑和心理学之间的相互关系存在着不同的看法。一些早期的逻辑学家，比如约翰·斯图亚特·密尔，就认为人类的心理学和作为推理的艺术和科学的逻辑学之间有很

密切的关系。与之相反，现代逻辑的创始人，古特罗伯·弗雷格和查理士·皮尔斯就强调他们的工作是远离心理学的。目前，关于形式逻辑和心理学的关系及相互之间的功过，至少可以区分出三种立场：

1. 形式逻辑是人类推理的一个重要部分。

2. 形式逻辑只是很间接地与人类的推理有关，但二者之间的距离无关紧要，因为逻辑在哲学和人工智能中的角色足以对是什么构成了优化的推理提供了数学上的分析。

3. 形式逻辑只是很间接地与人类的推理有关，所以认知科学应当寻求其他途径。

第一种立场为一部分心理学家所赞同，他们提供了人们使用像肯定式等逻辑规则的实验证据。第二种立场在偏重于形式方法的哲学家和人工智能研究者中很流行。第三种立场可能是目前心理学中的正统观点，但在哲学和人工智能中不太普遍。

心理学家里面最积极地推崇立场 1 的是马丁·布莱恩（Braine 1978）和朗斯·里普斯（Rips 1983，1986，1994）。里普斯（1986，第 279 页）列举了一批支持心理逻辑的心理学证据。心理逻辑的理论可以成功地预测出受试者对相当宽范围内的命题论证给出的有效判断。例如，人们能辨别出与肯定式形式上相同的有效论证，却拒绝像“如果 A 那么 C，C 成立，因此 A 成立”这样的论证形式。心理逻辑理论还能说明人们做逻辑判断时的反应时间，并有助于说明受试者报告有效决策的思考过程的陈述。

然而，其他类型的实验却让许多心理学家对心理逻辑持怀疑态度。最著名的是瓦森（Wason 1966）的选择性任务所使用的。受试者被告知他们所见到的纸牌一面有数字，另一面有字母。同时告诉他们一条规则，如“若一张纸牌的一面是 A，那么另一面为 4”。接着在被试者面前呈现 4 张纸牌，要求他们确定哪些纸牌需

要翻过来，以判定该规则是否成立。比如说，给被试者呈现的是如图2.1所示的四张纸牌。然后让他们确定哪些纸牌需要翻过来。大多数人能认识到有必要将A翻过来以检验另一面是否为4。这可以解释为是使用了肯定式，因为规则“如果A那么4”加上前提A，提示要检查另一面是否为4。然而，相当多的人却忽视了要检查7，没有意识到如果这张纸牌的背面为A，便与待检验的规则相抵触，因为如果另一面为A，则这一面应当为4。认识到带7的纸牌需要翻过来要求懂得否定式：如果A那么4，7意味着非4，这样如果规则成立那么另一面为非A。一些人竟糊涂到翻开B和4，实质上它们与检验规则是否正确毫不相关。

A B 4 7

图2.1 瓦森选择任务中的纸牌

这种实验的要点并不是表明受试者违犯形式逻辑的规则是愚蠢的。相反，该实验提示人们使用表征与计算完成这类推理任务的方式可能与形式逻辑的方式大不一样。随后的一些实验表明如果换成人们熟悉的例子，人们会毫不费力地完成与瓦森纸牌类似的任务。假如告诉受试者纸牌一面的信息是关于某人是否在酒吧间，面另一面的数字则代表该人的年龄，等待检验的规则是“如果一个人到酒吧去，那么他或她一定年满21周岁”，然后要求被试者回答哪些纸牌需要翻过来以确定该规则是否正确，比如说待选的纸牌有“在酒吧”、“不在酒吧”、“23”和“18”。与用字母和数字来表达的抽象问题不同的是，大多数人能认识到有必要翻开的不仅有“在酒吧”，以检查该人的年龄，还有“18”的纸牌，以确定该人不在酒吧里。

程和赫尔约克（Cheng和Holyoak 1985）论证说人们完成这些任务时不是使用了心理逻辑，而是采用了**实用推理程式**。例如，一个默认的模式是“如果一个人要做X，那么他必须先满足条件

Y”。这样，人们用具体的酒吧和年龄完成任务的情况比用抽象的字母和数字要好得多的原因，就在于这一默认的模式可以自然地运用于前者。关于规则和程式的心理学应用的进一步的讨论在第三章和第四章中介绍。

对心理逻辑观点最持久的批评来自心理学家菲利普·约翰逊-拉依德（Johnson-Laird 1983)。约翰逊-拉依德和拜恩（Johnson-Laird 和 Byrne 1991）提出演绎推理既不是使用形式化的逻辑规则，也不是有具体内容的规则或程式，而是用**心理模型**，这是在结构上与所表征的东西相对应的心理表征。他们认为当人们解释一个条件语句时，如“如果一张纸牌的一面是 A，那么它的另一面是 4”，先建构了一个心理表征：

[A]　　4

这里“[A]”代表一个模型，表示纸牌上有 A，而“4”加在这个模型上表示另一面有 4。约翰逊-拉依德和拜恩认为人们只考虑那些以明晰的方式在他们的规则模型中表征出来的纸牌，以此来解释要执行选择任务时多数人的表现，人们翻开纸牌 A 是因为它在他们建构的模型中得到了表征，而忽略了要翻开 7 是因为它未得到表征。借助于量词“所有的”和“一些”，心理模型理论还可以应用于其他形式的推理中。从形式逻辑的角度看，使用量词进行推理首先是运用诸如全称例示（上文曾经提到过）等推理规则移去量词，然后运用命题推理规则如肯定式来进行推理，最后再使用加置的推理规则重置上量词。看下面的简单例子：

所有的足球运动员都强壮。
强壮的人可以举起重物。
所以，所有的足球运动员都可以举起重物。

在逻辑上，通过例示将推理变成无量词的陈述是有效的推理形式，如“如果 x 是一名足球运动员，那么 x 强壮”和“如果 x 强壮，那么 x 可以举起重物”。命题逻辑可以得出“如果 x 是一名足球运动员，那么 x 可以举起重物”，再通过归纳推理原则进行概括使之对任何的 x 成立。与此不同，约翰逊-拉依德强调人们事实上使用的是模型而不是抽象的形式，可以建立这样的模型：

足球运动员　强壮　举起重物

在这样的模型里，没有举不起重物的足球运动员，所以所有的足球运动员都能举起重物这一结论也就水到渠成了。更复杂一些的推理形式带有“所有的”、“一些”和“非”的混合使用，这就需要更复杂一些的模型。约翰逊-拉依德认为人们对付这些不同种类的模型感到相对困难，正好对应于所建立不同类型的模型的复杂性。里普斯（Rips 1994）以及欧布莱恩、布雷恩和杨（O'Brien，Braine 和 Yang 1994）则回应说从心理逻辑上说明演绎推理的心理学证据来得比心理模型要好。

正如约翰逊-拉依德对形式逻辑与人类的演绎推理相关这一论点的挑战，一些心理学家通过实验也提出人类的归纳推理很可能与概率论没多大关系。例如，一个合取式的概率小于或等于其合取项的概率，即 $P(p\&q) \leqslant P(p)$。特维斯基和卡勒曼（Tversky 和 Kahneman 1983）的实验表明人们的推理经常违犯这一规则。假如，你被告知弗兰克喜欢阅读严肃文学作品，爱看外国电影，而且喜欢谈论世界政治，然后让你估计弗兰克受过大学教育、弗兰克是一名木匠以及弗兰克是一名受过大学教育的木匠三种情况各自的概率。受试者普遍认为弗兰克受过大学教育的可能性比他是一名木匠的可能性要大，这一点并不令人奇怪，但受试者通常却违反概率理论而判断弗兰克是一名受过大学教育的木匠的可能性比他是一名木匠的可能性要大。当人们面对这些例子时，他们似

乎使用了一种匹配过程来对关于某人的描述与他们关于某类人物的原型（如大学生和木匠）的吻合程度进行判断（参见第四章）。不少其他例子也提示人们的归纳推理似乎依据的不是概率理论的形式规则（Kahneman，Slovic 和 Kleinbolting 1991）。

一种可能的解释是心理逻辑可以对某些专门类型的人类推理（如运用肯定式）给出恰当的说明，而对解释那些复杂的人类推理（如引人了“所有的”和“一些”）则需要借助于一些更具体的表征（如心理模型）。至少可以说，逻辑不是理解人类思维的唯一途径，在后面的章节里我们会讨论其他的选择。当然，对心理学不感兴趣的哲学家和人工智能研究者可以坚持人类在思维时是否使用了逻辑对发展关于人类和其他智能系统**应当**如何思维的形式逻辑模型来说并不重要。但人类的智能和我们所要创造的机器智能很有可能依据的是与逻辑所提供的很不一样的表征结构和计算过程，对此他们要承担风险。

神经学上的合理性

目前对形式逻辑在神经学上的合理性还一无所知。我们对大脑的了解还太有限，以至于我们无法了解神经元是否使用形式化表征或者完成某种类似肯定式的过程。神经元之间的突触联接倒有点像下述的微型推理模式：如果神经元 1 被激活，那么神经元 2 也被激活。神经元 1 被激活了，所以神经元 2 也被激活了。但显然单个神经元不可能表征整个命题，而神经元组群怎样进行推理却不得而知。初看起来，第七章里要讨论的联接主义模型比逻辑有更多的合理性。但我们不能排除这样的可能性，即人的神经网络所做的在逻辑上正好是进行推理。这种表现在受过逻辑训练的人身上肯定发生，即便对一般人来说逻辑来得并不自然。

实践上的可应用性

对于深入了解人类的学习而言，认知科学的逻辑途径在教育上没能发挥出多大用处。皮亚杰和英赫尔德（Piaget 和 Inhelder 1969）曾试图将人类认知发展的一些原则置于逻辑范畴之上。但有关命题逻辑在发展阶段的角色的主张却未能成为现代教育理论的一个部分。不过，从教育的另一个角度来看，逻辑对于启发人们应该怎样更好地推理还是有用的。形式逻辑和批判思维的课程受到欢迎，其原因是人们需要改进推理能力。形式演绎逻辑和概率理论确实为描述一些类型的思维活动提供了有用的工具。

据戴姆和列维特（Dym 和 Levitt 1991）称，工程设计通常涉及到对一些要求的满足，而这些要求可由逻辑陈述来表达。例如，一条有关结构的规程可以这样表述："如果一根梁柱要能被支撑住，那么它的深度应比它的净跨度的1/30要长。"PROLOG 是使用逻辑表达和演绎技术的一种编程语言，可运用到设计需满足一定的物理约束和法律约束的建筑物这类设计问题上。虽说逻辑是人工智能理论家喜好的工具，实际的智能系统却倾向于使用后面将要谈到的规则、案例和神经网络等方面的技术。

小　　结

形式逻辑为了解表征与计算提供了一些有力的工具。命题演算和谓词演算可用于表达多种复杂的知识，而许多推理可以根据逻辑演绎的规则（比如肯定式）来得到理解。认知科学的逻辑途径的解释程式是：

解释目标

为什么人们会作出推理？

解释模式

人们具有类似于谓词逻辑里的语句的心理表征。

人们具有在这些语句上进行操作的演绎和归纳的程序。

演绎和归纳程序，运用到语句上产生出推理。

不过，逻辑是否能为认知科学的表征和计算提供核心概念，这一点还不能确定，因为解释人类的思维可能需要更有效的、在心理学上更自然的计算方法。

讨论题

1. 你所知道的什么事用形式逻辑来表达很困难？
2. 人们是合逻辑的吗？他们应当合逻辑吗？
3. 演绎是人类思维的核心方式吗？人们是怎样进行演绎的？
4. 非演绎推理符合概率原则吗？
5. 自然语言以逻辑为基础吗？

进一步的推荐读物

关于逻辑的历史，见 Prior 1967。有许多不错的导论性逻辑教材，例如 Copi 1979。Pollock 1989 从基于形式逻辑的计算角度来讨论哲学问题。从逻辑的途径研究人工智能在 Genesereth 和 Nilsson 1987 进行了详细的阐释。Rips 1994 对人类的演绎思维的逻辑的途径进行了深人的探讨和辩护。

备注

形式逻辑关心的不只是句法，即在本章中所涉及到的语句的结构，还关心语义学，即语句的真值条件。例如，合取 $p\&q$ 只有在 p 为真且 q 也为真的

情况下才为真，在其他情况下则为假。

人工智能领域的大多数研究基于逻辑的规划的学者并不使用一整套的逻辑推理规则，而是使用一种基于**归结原理**的简单推理规则的推理程序(Genesereth 和 Nilsson 1987)。归结原理本身较为复杂，在此不作详细讨论，它是将形式表达转换成一套简化的谓词演算，并使用一种强有力的算子来推导演绎的结果。法克思和尼尔逊 (Fikes 和 Nilsson 1971) 将逻辑演绎用到一个规划系统 STRIPS 中，该系统用于机器人和其他应用中。

贝耶斯定理用符号表示为：

$$P(h/e) = \{P(h) * P(e/h)\}/P(e)。$$

第三章 规　　则

规则是这样的如果-那么结构：**如果**你通过了40门文科方面的课程，**那么**你将获得B.A.（文科学士）。这些结构与上一章讨论的条件句非常相似，它们却具有不同的表征和计算性质。虽则大多数基于逻辑的计算模型并不打算作为人类的认知模型，基于规则的模型却从一开始就具有心理学目标。第一个人工智能的程序是阿兰·纽威尔、克列夫·肖和赫伯特·西蒙（Newell，Shaw和Simon 1958）编写的"逻辑理论家"。这个程序是于1956年在一台原始的计算机上写的，用来证明形式逻辑中的定理。它的研制不仅是要成为一个在数学上是足够复杂的智能系统，还要成为人类怎样证明逻辑定理的一个模型。除逻辑推理的规则外，"逻辑理论家"还含有有效寻找证据的策略性规则。不久，"逻辑理论家"就被推广到第一个试图理解人类思维的大框架之中：GPS——通用问题求解器（Newell和Simon 1972）。GPS采用规则来模拟人类解决各种类型的问题，如本章随后将要提到的密码算法问题。

自GPS之后，两种基于规则的认知系统在认知科学中影响较大，它们在对人类认知的研究中得到了广泛的应用。约翰·安德森的ACT系统（Anderson 1983，1993）在心理学方面的应用较广。近期，阿兰·纽威尔与约翰·拉依德及保罗·罗森勃卢姆合作开发了SOAR，这是一个有着很多技术上和心理学方面应用的强有力的基于规则的程序（Newell，1990；Rosenbloom，Laird和Newell，1993）。

本章的重点不是在细节上介绍这些系统，而是要讨论是什么使得规则在计算和心理学方面如此有力。后面的章节，则会提供看待认知的其它视角，从而表明规则并没有揭示出人类思维的全貌。

表征力

尽管规则是一种非常简单的结构，只有一个**如果**部分（有时称为**条件**）和一个**那么**部分（称为**行动**），它们却能用来表达多种不同类型的知识。首先，它们可用来表达关于这个世界的一般性信息，如学生负担过重：**如果** *x* 是一名学生，**那么** *x* 负担过重。其次，它们可用于表达如何完成一些事情的信息，比如：**如果**你尽早登记，**那么**你就可以选上你想要学的课程。第三，规则可以表达语言使用上的规定性，例如：**如果**一个英语语句有复数主语，**那么**它必须有一个复数型的谓语。第四，上一章里我们提到的推理规则如肯定式，可以改写成：**如果**你有一个如果-那么规则且如果部分为真，**那么**它的那么部分亦为真。正如这个例子所示，规则可以具有多重条件（**如果**部分中的多个子句），而且可以含有多重行动：**如果**及早登记，**那么**你可以选上你想要学的课程，并且还可以少排队。

初看起来，基于规则的系统在认知科学中如此重要，这似乎让人感到惊奇，因为规则远不如形式逻辑在表征上来得优雅。逻辑提供了一种标准化的方案来表达关系及诸如“与”、“或”和“非”等基本操作，而这些在基于规则的系统中是由各种非标准化的方式来实现的。但规则系统的开发者却乐于牺牲某些逻辑系统在表征上的严密性，以换取更强的计算力。由此而来的一个优势是规则没有必要解释为普遍为真。（对所有的 *x*）（是学生（*x*）→负担过重（*x*））这一逻辑概括必须解释为每一个学生都负担过重。而规则**如果** *x* 是一名学生，**那么** *x* 负担过重则可以解释为一个默认（default），即作为一个可以允许例外的大致的概括。我们还可以有另一条规则，**如果** *x* 是一名学生并且 *x* 只修选容易的课程，**那么** *x* 并不负担过重，这两条规则可以在同一个系统中共存，故而其结果不一定是某一名学生负担过重同时又不负担过重这样

的矛盾性结论，这是因为基于规则的系统的计算操作可以确保只运用其中较为恰当的一条规则。

与逻辑不同，基于规则的系统可以方便地表达关于该做什么的策略性信息。规则通常含有表示目标的行动，例如：**如果**你想回家度周末并且有路费，**那么**你可以乘坐公共汽车。这种有关目标的信念使得基于规则的问题求解者集中解决当前的任务，因此，尽管规则系统中的规则也许不具备形式逻辑的全部表征能力，对它们的表述却可以采用能增强计算力和心理学上的合理性的方式。

计 算 力

问题求解

在基于逻辑的系统中，思维的基本操作是逻辑演绎，而对基于规则的系统来说，思维的基本操作是**搜索**。当你有一个问题需要解决时，例如为一门课程撰写一篇论文，你会有一个供你选择的可能性**空间**。这个空间包括可供你选择的题目，你可以参阅的图书馆的文献资料，以及在你实际撰写过程中可以借助的工具。完成这一任务就要求你在这个可能性空间中进行搜索，以找到一条从你的当前状态（要写论文）到达目标状态（完成一篇使你获得高分的论文）的途径。基于规则的系统可以有效地完成这类搜索。对复杂的问题，不可能穷尽地搜遍整个可能性空间以求得最佳解决方案。例如，假设你要穿戴 4 件不同的服饰（衬衣、裤子等等），而每种服饰你都有 10 件，（10 件衬衣，等等），这样你每天可能的穿戴便有 1 万种（10^4）不同的组合方式，没有人会有时间或兴趣来考虑所有的这些可能。实际上，人们依靠的是**启发式方法**（heuristics），即依靠经验性规则以求得满意的解决，而毋须考虑所有的可能。“棕色的鞋子配棕色的裤子而别配黑色的裤子”，这样一条启发式信息，就有助于对穿戴的规划问题提供一个有效的

解决。问题求解、学习和语言的使用均可以看成是在一个复杂的可能性空间里进行基于规则的启发式搜索。

心理学家对长时记忆和短时记忆作出了重要的区分，前者是心智中长久的信息存贮，而后者是选出的小量可供即时加工处理的信息。从基于规则的角度看，在你的长时记忆存有许多规则，但仅有少量的规则和事实激活在你的短时记忆中，以备当前使用。你可能把你母亲的生日存在了长时记忆中，而读到这个句子时使你意识到了你母亲的生日，因为它已在你的短时记忆中被激活了。

计算机科学家和心理学家对串行处理和并行处理作出了重要区分，前者是指一次只做一步操作，而后者是在同时做多项操作。基于规则的推理既可以是串行的，每次你只使用一条规则；也可以是并行的，多条规则同时被使用。有意识的思维一般是串行的，我们会注意自己每次只做了一步推理，但这些推理也可能依赖于多条规则的同时使用，对此我们可能意识不到。第九章会讨论意识在思维中的角色。

规划 很多学生就学的大学不在家乡，所以他们经常会面临的一个问题是在周末或期末的时候怎样返家。从学校到家的途径可以由一系列规则来表达，例如：

如果你经由高速公路 1，**那么**你可从大学所在地返回家乡所在地。

如果你经由乡间公路，**那么**你可从大学所在地到达高速公路。

如果你从学校走主大街，**那么**你可从学校上乡间公路。

如果你从公共汽车站乘公共汽车，**那么**你可从大学所在地返回家乡所在地。

如果你从学校乘公共汽车到汽车站，**那么**你可到达公共汽车站。

其它的可能性也是存在的，诸如乘火车或者搭乘便车。要解决如何返家度周末这一问题的学生可以在这个可能性空间里进行搜索（去公共汽车站、上高速公路），并把它们组成一个计划，从而到达其目的地。

既可以从**正向**也可以从**反向**运用规则进行推理。进行反向推理，你可以想到“要回家便要经高速公路，而这就要求先经乡间公路，进而要求先经主大街，而这又要求有一辆小汽车”。目标是回家，而这一计划可以由考虑一系列的子目标来构造，比如要先到达高速公路。正向推理，则可以使用近似于肯定式的推理，经由主大街可达乡间公路，进而抵达高速公路。正向推理和反向推理，都是试图找到一系列规则使你从起始点到达目标，只是它们在搜索策略上不同而已。

另一种可能的推理策略是**双向式**搜索，由从起始点向前搜索和从目标向回搜索组合而成。虽说许多规划问题可以从基于规则进行推理的方式来理解，但当存在很多潜在相关的规则时，你不得不选择哪些规则用在求解问题的关键之处，以这种方式来进行规划就变得比较困难些。基于规则的问题求解在很多地方看起来像是逻辑演绎，不同的地方在于它将更多的注意力放到在适当的时机运用适当的规则的策略之上。

学生在课堂上遇到的规划问题也是这种情况。一道数学文字题给定你一些信息，然后要求你计算出答案。例如，告诉你从学校所在地到家庭所在地要花 75 分钟，距离为 65 英里（100 公里），要求计算出平均速度。规则，体现在数学运算中，教你怎样正向地从给定的信息求得从中可推导出来的答案。当然，通常情况下，反向地从目标——欲求得的答案——朝着给定的初始信息求解会更有效。不管怎样，都是要求你找出一系列规则以提供从起始到达目标的途径。然而，并非所有的规划都是基于规则的。在后面的章节里我们会看到程式和类比如何有助于解决规划问题。

决策 虽然规则对于寻找规划很有帮助，对于在多个规划中进行决策却不是太有用。学生也许能够使用规则构思出两条回家度周末的不同路线。但这不足以对选用哪一种方案提供指导。驾车、乘坐公共汽车和乘火车都可以使你回到家，而你选择哪条路则需要你在不同的目标之间进行更复杂一些的平衡，如节省路费、节约时间和避开拥挤。因此，对决策制定来说，基于规则的推理需要由其它的处理过程加以补充，如第二章提到的期望值计算，或是第七章讨论的并行约束满足。

解释 在第二章中我们已经看到，解释通常被看作是一种演绎过程，在这里规则与逻辑演绎一样可以胜任。某些类型的假设形成可以描述为根据规则对解释进行的搜索。假如你想要选修一门课程而它已经被选满了。各种规则可能是这样：

如果一门课程是很多专业的必修课，**那么**它很快就会被选满。
如果一门课程的教师很受欢迎，**那么**它很快就会被选满。

如果知道了这门课的教师很受欢迎，加上上述的第二条规则，这就可以解释为什么当你去登记选修那门课时已经满员了。即便你不能确定这门课有一位很受欢迎的教师或者是许多专业的必修课，你也可以猜测这些情况可能是真的（参见下面对逆推式学习的讨论）。这样，如果规则导出的结果使你能从你已有的信息得出你要解释的东西，解决解释问题就可以根据基于规则的推理来理解。

学　习

许多种重要的学习都可以自然地理解为对规则的获取、修正和运用。某些规则可能是先天性的，由我们出生时的生理结构所决定。**如果**有物体朝你的眼睛飞来，**那么**你会眨眼，这样一条物

理性规则就不会是人或者其它生物通过学习获得的。下一节要谈到一些认知科学家认为许多语言规则是先天性的，对此有较大的争议。但没有人会宣称怎样登记选修大学课程的规则是先天的，那他们又是怎样获知的呢？

与上一章提到的逻辑陈述相似，规则可以通过**归纳概括**来学习，由一条规则来概括一些事例。有些时候规则要求有很多事例来支持：你不应该仅从一门工科课程难学就概括出所有的工科课程都很难，或是从一门哲学课得出所有的哲学课都有趣的结论。但同学们会逐渐从经验中获取这样的规则：**如果** x 是一门编程课，**那么** x 会很耗费时间或者**如果**你想选上很受欢迎的课程，**那么**你应该尽早登记。

在归纳概括中，规则是通过事例形成的；但规则同样也可以由其它规则来形成，在认知模型 SOAR 中有这样的一个处理过程叫**组块**（chunking），而在 ACT 模型中叫**合成**（Composition）。假如你已经使用了大量的规则设计了一个从学校返家的方案，并且已找到了一系列关于怎样从学校到乡间公路再经高速公路抵家的规则。在下一次你再回家时你便不需要再从头来一遍完全的搜索。你可以把这一系列规则组合成为一条普遍性的规则：**如果**你要从学校回家，**那么**驾车。同样，你第一次安排课程表时，你可能要做大量复杂的搜索以设计一个好的方案，但后来你有了经验，便可使用一条更高层次的规则：**如果**你要安排一个好的课程表，**那么**把你的课程安排得紧凑些而不要分布到一周的五天。在心理学合理性一节里我们将会看到将规则组块的计算过程已在多种有关人类学习的模型中得到了应用。

由规则生成规则的另一种办法是**特殊化**，对一条已有的规则进行修正以处理一些特殊情况。如果星期五下午返家时由于交通堵塞可能会很费时，经验可能会让你得出一条专用规则：**如果**你要从学校返家并且是在星期五下午，而且你有急事，**那么**就别驾车。

如同我们在上一章对逻辑的讨论一样，规则亦可用于逆推式

学习。假如你的一位朋友生气而且沮丧，很自然地，你会试图去解释是什么引他（她）不快。如果你已从归纳概括得到这样一条规则：**如果**一名学生考得不好，**那么**他（她）会生气而且沮丧。这样你会猜测你的朋友可能考试成绩很糟，从而对他（她）的不快给出一个可能的解释。这就是逆推式推理，规则被反向地使用以便对所发生的情况给出一种可能的解释。显然，这种推理是有风险的，因为对于你朋友的心理状况可能还有一种更好的解释，比如有一种解释是基于规则：**如果**某人遭同伴拒绝，**那么**他（她）会生气而且沮丧。挑出最好的解释需要一种复杂的推理，这在第七章讨论，但规则对于形成诸如你的朋友考砸了这类的假设还是非常有用的。由此可见逆推式推理与基于规则的推理能自然地结合（Thagard 1988），虽然我们会看到其它类型的表征也同样能支持它。

如果对每条规则增加一个代表其适用性或合理性的数量值，规则还可用于描述那些缓慢增强的学习过程。对一条规则的使用越是成功，它就越是合理和有用。例如，如果一名学生每次都采用了规则：**如果**你要从学校返家，**那么**驾车，这条规则就变得越来越强，在今后就越有可能使用这条规则。总之，规则既可以从事例中产生，也可以来自于其它规则，可以逆推式地运用，还可以在其效果的基础上进行量化评价。

语　言

在50年代认知革命以前，语言曾被普遍地认为是由通过联想学习的行为构成的。通过对词语对的重复性经验，人们会逐渐习惯并期待同时使用它们。在1957年出版的《句法结构》一书中，语言学家诺姆·乔姆斯基提出了一种大异其趣的语言观。乔姆斯基认为行为主义者的学习模型无法说明语言的生成性，即人们能够生成和理解数不清的语句。你也许从未接触过“她骑着一头紫色的骆驼去学校”这样一个句子，但理解它却没有任何问题。

按照乔姆斯基的学说，我们表达和理解语言的能力依赖于我们能处理由规则组成的一整套复杂语法，尽管我们没有有意识地认识到这些规则。例如，学习英语的儿童，对过去时的动词加上“ed”，并未意识到他们使用了一条规则：**如果**你要使用一个动词来描述过去，**那么**在此动词的后面加上“ed”。众所周知，5岁以下儿童过泛地使用了这一规则，用“goed”和“bringed”而不是使用这些动词的不规则形式。阿马简等人（Akmajian 等 1995）介绍了规则在语言的几个不同方面的应用。例如，说英语的人知道怎样使动词加上“er”而变成名词，如把“write”变成“writer”。我们还知道诸如怎样拼读复数名词的语言学规则：比较一下“cats”/“huts”与“dogs”/“hugs”中“s”的不同发音。句法规则使我们能够将陈述句变成疑问句，如把助动词“am”移至句首便将“I am happy”变成了“Am I happy?”在很多问题上，乔姆斯基富有影响力的观点都引起争议。在第七章，我们将讨论联接主义者关于语言的看法，即语言不是由规则构成，而是由简单处理单元之间联接权重所表征的更松散的联接构成的。撇开我们关于语言的知识是否由规则表征这一问题，另一个问题是这种知识是习得的还是天赋的。乔姆斯基仍坚持强调对每一个人来说一种天赋的普遍语法是与生俱来的。他放弃了一些早期的观点，即儿童是通过形成关于哪些规则运用到各自的语言中的假设，从而逆推式地获得使用语言的能力（Chomsky 1972）。他目前认为儿童仅需通过认知语言所使用的一个有限的可能性集合即可自动地学会一门语言（Chomsky 1988）。所有人类的语言都具有名词、动词、形容词及前置介词或后置介词，但有的语言如日语就没有像英语中“the”和“a”这样的冠词，所以学日语的儿童必定是以一种与学英语的孩子所不同的方式来实现普遍语法的。

尽管乔姆斯基关于语言由天赋规则构成的观点对语言学和相关领域产生了巨大的影响，我们仍会在后面的章节讨论其它看待语言的方式。

心理学上的合理性

在本书介绍的所有计算-表征理论中，哪一种在心理学上得到最多的应用？答案是：基于规则的系统。我无法在此对所有这些心理学应用给出全面的评价，但我会提供运用基于规则的系统来说明人类思维的一些典型方式的例子。

纽威尔（Newell 1990）向我们展示了 SOAR，一个复杂的基于规则的系统，如何能应用到相当多有趣的心理学现象上。例如，他描述了 SOAR 怎样解决一些密码算术问题，即用数字替换字母的字谜（也可参阅 Newell 和 Simon 1972）。一个字谜是 DONALD +GERALD=ROBERT，每个字母必须由 0 到 9 之间的一个数字替换，以使得方程式成立。在解决这个问题时规则十分有用，为便于使用我们常用的加法算法可将此题更清楚地表达为：

DONALD
GERALD
ROBERT

一个人怎样着手解这个字谜呢？你可能会注意到左边第二列 O 加上 E 得到 O，这可能使你记起规则：**如果** 0（零）加上一个数字，**那么**该数字不变。这条规则提示 E 是 0。然后，在第四列你会看到 A+A=0，A 应当为 5。但是沿着这条线索推下去很可能使你陷入困境，因为这假定了在 L 加 L 得到 R 或者 N 加 R 得到 B 的运算中没有进位。因此，R 必须是 L 的 2 倍，但又不能太大，这样在第三列加上 N 而不至于产生进位。在我们熟悉的加法运算中进位涉及到规则：**如果**数字相加超过 10，**那么**写下这个数的个位，并且把 1 进到左边一列。这条规则表明 E 的值还有另一种可能性：在 N+R=B 的这列有进位的情况下的，如果 E=9，那么 O+E=O。从这一点出发，使用相加和进位的规则，再加上一些别的知识，比如：**如果**一个数字在一个数的开头，**那么**这个数字就

不是 0，便可使你得到这个字谜的答案（当然你得费点劲）。SOAR 通过设置各种算子，推算各个字母的数值，并检查它们之间是否一致，不发生冲突，从而反映出这一过程的各个方面。

SOAR 还可用于模拟其它一些高层推理活动，如确定从“一些弓箭手不是投球手”和“所有的划桨手都是投球手”，可以导出什么结论来。SOAR 没有采用第二章中讨论过的两种演绎方式：心理逻辑和心理模型，而是在一个可能的推导空间中进行搜索，逐步形成结论：“一些弓箭手是划桨手”。这在逻辑上是不正确的，但从说明演绎思维的认知角度看能够模拟人类的推理过程，包括人们有时如何会出错。

纽威尔还运用 SOAR 来说明人类学习的许多方面，特别是可以说明练习的幂方定律，即学习的速率会随着所学内容的增多而降低。这一规律体现在许多的学习任务中，诸如打字以及书写经转置的字母，经转置后的字母就像字母映在镜子里一样。SOAR 中的组块功能可以对为什么学习的速度随着你对某一任务经验的增多而降低的原因作出解释。在任务练习的开始阶段，你可以迅速地构造很多的组块，而随着组块的建造，行为的速度也增加了。例如，开始练习打字时在速度和准确性上都有一个较大的飞跃。但随着更高层组块的形成，对它们的使用会越来越少，因为它们适用的情境是很稀有的。因而随着练习次数的增多，一个基于规则的系统的学习速率会逐渐降低，在这一点上与人的情况是相似的。

霍兰德等人（Holland 等 1986）采用基于规则的系统来说明多种不同类型的学习。例如，对老鼠来说，它们要学会怎样避免遭撞击，这可以由假设它们使用规则的学习是通过调整所使用规则的不同强度来加以解释。每一次老鼠按下杠杆就得到食物，规则**如果**杠杆，**那么**食物就得到了增加。相反，按下杠杆就遭到撞击会产生出一条矛盾的规则，**如果**杠杆，**那么**撞击，从而使老鼠停止按动杠杆。人们应付物质世界和社会环境的能力和极限同样可以由规则来加以理解。例如，一旦人们了解了一种社会角色的原

型，他们会倾向于过于普遍地运用这一原型。

克劳利和西格勒（Crowley 和 Siegler 1993）揭示了儿童玩三连棋——一种力求在一列中得到 3 个×或 3 个 O 的简单游戏——能力上的差别怎样可以通过他们对规则的习取而得到理解。儿童需要掌握有关走棋子的规则，还要了解在适当的时候使用恰当的规则的策略性知识。这里是一部分规则：

获胜：**如果**在棋盘上的一行、列或对角线上，有我方两个棋子并有一个空格，**那么**在空格上放入棋子就获胜。

阻击：**如果**在一行、列或对角线上有对方两棋子并有一个空格，**那么**就在空格上放入棋子以阻止对方获胜。

占中：**如果**棋盘的中央空着，**那么**在棋盘中央放入棋子。

占空角：**如果**棋盘上有空角，**那么**在棋盘空角上放入棋子。

图 3.1 显示了上述四条规则都采用了的一个棋局。儿童玩三连棋的能力随着他们掌握的规则的增多以及他们对规则使用的优先顺序的认识而增强。例如，许多学龄前儿童就没有掌握阻击规则，而许多掌握了这一规则的孩子却会在有机会获胜的时候还使用该规则。

基于规则的系统还适用于说明对语言的学习和使用。安德森（Anderson 1983）用下述规则描述了人们关于英语的知识（以最简化的形式）：

如果目标是表达一个形式（关系、主体、客体）的有意义的结构，**那么**分解为子目标：

1. 描述主体
2. 描述关系
3. 描述客体

其它的规则则表明怎样完成子目标以得一个完整的句子，比如"那姑娘投了一个球"，这个句子用以描述主体对客体做了什么。安德森（Anderson 1993）介绍了他的基于规则的 ACT 系统的大量的应用，诸如求解几何题和计算机编程。有很多例子能使得基于规则系统的性能与人类的行为相吻合。

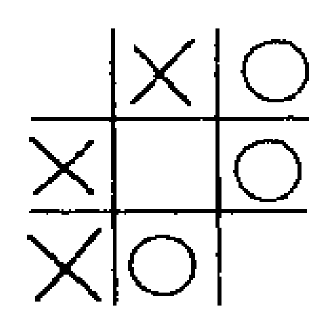


图 3.1　规则在三连棋中的应用

轮到×走棋，适用于四条不同规则的**如果**部分：获胜（在左上角走棋）；阻击（右下角）；占中（在中央走棋）；以及占空角（在任一个空角走棋）。经授权选自 Crowley 和 Siegler 1993，第 537 页。

神经学上的合理性

在规则与由突触相联接的神经元之间有一点粗略的类比关系：**如果**一个神经元被激活了，**那么**它引起与之相联的神经元被激活。但这种相似性是很表面化的，事实上对规则怎样由大脑来实现可以说知之甚少。安德森（Anderson 1993）勾画了 ACT 由神经元来实现的一种可能的轮廓，并介绍了一些已经由人工神经元网络（见第七章）实现的简单的基于规则的系统。早期的基于规则的系统都是串行的，近期的基于规则系统如 SOAR 可以同步地激活多条规则，以模拟大脑的并行活动。不过，大多数基于规则系统的发展是以认知模型独立于神经学上的考虑为假设的。这一假设受到了联接主义的挑战，我们会在第七章加以讨论。

实践上的可应用性

如果说我们所学习的东西是由规则所构成的，那么教育就应关注帮助儿童和学生更好地掌握规则。安德森(Anderson 1993)介绍了他的 ACT 基于规则系统在教学方面的大量应用，包括理解人们怎样学习计算机编程、文本编辑和证明几何题。基于规则的系统不仅用来模拟学习者的表现，还用于建造计算机辅助教学系统以帮助人们学习。

工程上和其它领域的设计工作也可由规则来理解。纽威尔(Newell 1990) 介绍了 SOAR 的一个版本，通过使用算子来生成和测试程序规格从而在可能的算法空间中进行搜索来设计计算机算法。他和他的助手们还探讨了将计算机的使用者看成基于规则的系统，以便设计更便于人们使用的计算机 (Card，Moran 和 Newell 1983)。

应用于工业和政府部门的大多数专家系统是基于规则的系统，这是最早开发的一类应用型智能系统(Buchanan 和 Shortliffe 1984；Feigenbaum，McCorduck 和 Nii 1984)。很多领域的专门知识、从构组计算机系统到探测石油，都可以由规则来刻画。最近的一些基于规则的专家系统（以及其它类型的专家系统）的例子可以在 AAAI 出版社的《第六届人工智能新颖应用大会论文集》(包括前几届的论文集)中找到。朗格利和西蒙(Langley 和 Simon 1995) 提供了大量从案例中学习规则的计算机程序在工业中应用的例子，包括化工过程控制、信贷决策以及机械设备的故障诊断。

小　　结

相当多的人类知识可以很自然的用规则来加以描述，并且很多类型的思维活动如规划可以由基于规则的系统来模拟。这里所

用的解释程式是：

解释目标

为什么人们是有某种特定类型的智能行为？

解释模式

人们具有心理规则

人们具有使用这些规则在一个可能的解答的空间进行搜索的程序，以及形成新规则的程序。

使用和形成规则的程序产生出行为。

基于规则的计算模型能对相当广泛的心理学实验提供详尽的模拟，从密码数字的问题求解到语言的运用和技能的获取。基于规则的系统还具有实践上的重要性，可以对改进学习和开发智能机器系统提供启示。

讨论题

1. 你所具有的哪个领域的知识易于用规则来描述？
2. 你所具有的哪个领域的知识难于用规则来描述？
3. 基于规则的方式与上一章介绍的逻辑的方式有何不同？
4. 大脑可能以什么方式来实现规则？
5. 有关语言的知识是天赋的还是习得的？

进一步的推荐读物

将心智视为基于规则系统的经典读物有 Newell 和 Simon 1972，Newell 1990 以及 Anderson 1983，1993。Holland 等 1986 讨论了基于规则系统的多种类型的学习。Smith，Langston 和 Nisbett 1992 探讨了在推理中使用规则

的情况；亦可参见 Nisbett 1993。Pinker 1994 是对乔姆斯基语言观的一个引人入胜的辩护，其中谈到了规则对语言的重要性。

注　释

在逻辑上，一条规则或条件句的如果部分被称为前件，而那么部分被称为后件。在 AI 中，规则通常被叫作产生式。

认知科学的其它非规则研究途径同样也可以使用搜索一词。但搜索一词主要是与基于规则的系统联系在一起的。搜索隐喻能较好地处理那些能良好定义的问题，即状态和算子都能清楚确定，但对那些涉及学习新的表征和算子的任务的问题就不太有力了。

有关练习的幂方定律可以更技术化地表述为："如果以练习次数的对数为横坐标，以完成任务的反应时为纵坐标，则其结果为一条向下倾斜的直线"。

第四章 概 念

当学生们开始适应他们的校园生活时，他们不仅会学到许多新的规则，他们还需要掌握一些新的概念。许多新的学籍管理的概念是必须了解的，如**专业**、**注册**和**辅修**。学生们还会很快学会一些关于课程的新概念，比如用**小鸟**或**香肠**或是**蛋糕**来形容一门特别容易的课程。关于校园社会生活的知识也会迅速的增加，学生们会掌握一些描述不同类型的同学的概念，如**电脑迷**、**健将**、**大师**（**艺术系学生**）**和牲口**（特别勤奋的学生）。上过像讨论班和大型报告会等不同类型的课的学生就必须修正他们在中学时所熟悉的关于**上课**的概念。

关于概念在知识中的角色的问题可以追溯到2000多年前的古希腊哲学家柏拉图，他提出了"什么是正义?"和"什么是知识?"这样的问题，并揭示出像**正义**和**知识**这样的概念是很难定义的。柏拉图认为有关这类概念的知识是天赋的，而教育的作用是唤起我们对这些概念的本质的回忆。恰如乔姆斯基强调语言规则是天赋的一样，柏拉图和其后的莱布尼茨和笛卡尔等哲学家认为最重要的概念完全来自于心灵。

其他一些哲学家如洛克和休谟则坚持概念是通过感觉经验习得的。例如，你获得关于**狗**的概念，并不是仅仅通过思考狗是什么，而是通过接触各种各样的狗而获得。虽说杰里·福多（Fodor 1975），一位深受乔姆斯基影响的当代哲学家，坚持概念主要是天赋的，但当今大多数的认知科学家感兴趣的是从经验或者是其它概念而学习新概念的具体过程。

在本世纪70年代中期，随着研究者们引入了诸如"框架"、"程式"和"脚本"等术语来刻画对概念本质的新见解，对概念的

本质的心理学和计算上的兴趣空前地兴盛起来了。(某些类似的思想曾经由巴特里特（Bartlett 1932）和康德（Kant 1965）提出过。）在这一期间最具影响的一篇人工智能文章中，马文·明斯基（Minsky 1975）指出思维应当理解为对框架（frame）的应用而不是逻辑的演绎。在一项计算心理学的合作研究中，香克和阿倍尔逊（Schank 和 Abelson 1977）揭示出我们所具有的大量的社会性知识如何由脚本（script）来构成，脚本描述典型性的时序事件，比如走进餐馆用餐所发生的事情。与此同时，心理学家罗姆哈特（Rumelhart 1980）采用被称为程式（schema）的类似于概念的结构来描述知识，程式表征的不是像狗这样的概念的所谓实质，而是狗的典型特征。与之相似的是，哲学家哈拉里·普特南（Putnam1975）提出概念的意义应当视为其原型（Stereotype）而不是它的定义条件。在 80 年代，随着联接主义模型的发展，从计算的角度对概念的探讨呈现出一幅完全不同的面貌，这一点我们留待第七章来讨论。

表征力

多少次你会听到别人要求：“定义你的术语！”人们经常会这么说:“在我们定义智能这个词的意思以前我们无法谈论智能。”这一要求需要的是一个定义来提供准确的规则：**如果** *X* *拥有智能*，**那么** *X* *具有属性* *Y* 以及**如果** *X* *具有属性* *Y*，**那么** *X* *拥有智能*。然而正如柏拉图所发现的那样，诸如**正义**这样的概念，是很难找到这样的定义的。作为练习，请试着给**学院**、**大学**、**课程**或**家伙**这些概念找出能准确抓住它们含义的规则。实际上，像智能这样的一个概念只能在研究探索的后期才能得到定义，而不应在研究的起始。在数学之外的领域，我们甚至就不应当期望对概念给出完全准确的定义。

由框架、程式或脚本构造而成的概念是从对典型实体或情境

的表征来理解的，而不是根据严格的定义。例如，学生要了解关于**课程**的概念，课程由教师讲授指导，学生在课程结束时得到一个成绩。一门课程可以由一组槽（slot）来加以概括，每个槽里填上相应的信息，比如教师的姓名。

课　程

一种类型的：过程（一系列系统性的行动）

课程的种类：讲座课程，讨论班，等等。

教师：

教室：

授课时间：

要求：考试，论文，等等。

例子：哲学 100，数学 242，等等。

学生在报名选修一门课程时首先要做的事情之一是找出授课教师是谁，由此就填上一个重要的槽。**课程**概念也可以由一组规则诸如**如果** X 是一门课程，**那么** X 有一名教师来得到另一种表征，但在下一节里我们会看到有许多计算上的考虑使我们选择一组槽来表征一个概念。虽说一般的课程都有一名教师，但这不应看成是对课程的定义的一部分，因为有的课程会有不止一名教师，而有的课程如函授课可能就没有教师。

有的概念涉及到时间顺序，如考试这个概念：

考　试

一种类型的：课程要求

考试的种类：书面、口头、开卷、简答，等等。

地点：

顺序：

拿到考题

在答卷上写上姓名
回答问题
检查答案
交回答卷

同样，这一概念也可以看成一组规则，**如果**你参加一门考试，**那么**第一步先拿到考题，但从认知的角度看将概念视为可以作为整体使用的一套信息封装更为有用。

虽然概念中的槽通常可以转译为规则，但要注意槽所表达的不是普遍真理，只是一种典型的情况。槽的值有时被称为默认值(default)，一门课程的教师人数的默认值是一名。规则也可以理解为表达默认期望值而不是普遍真理，如在语句**如果** X 是一门课程，**那么** X 一般来说有一名教师。考试通常在校园里的教室进行的，但开卷考试是一个例外，并没有某个指定的教室，要求每个参加考试的人都到那儿去答卷。

概念能够以一些很重要的方式组织知识，这是基于规则的系统通常做不到的。请注意概念**课程**包括陈述一门课程是一类什么事物的槽以及具有哪些种类的课程的槽。这些类别关系就建立了一个概念的分层网络，如图 4.1 所示。一门讨论课是一种类型的

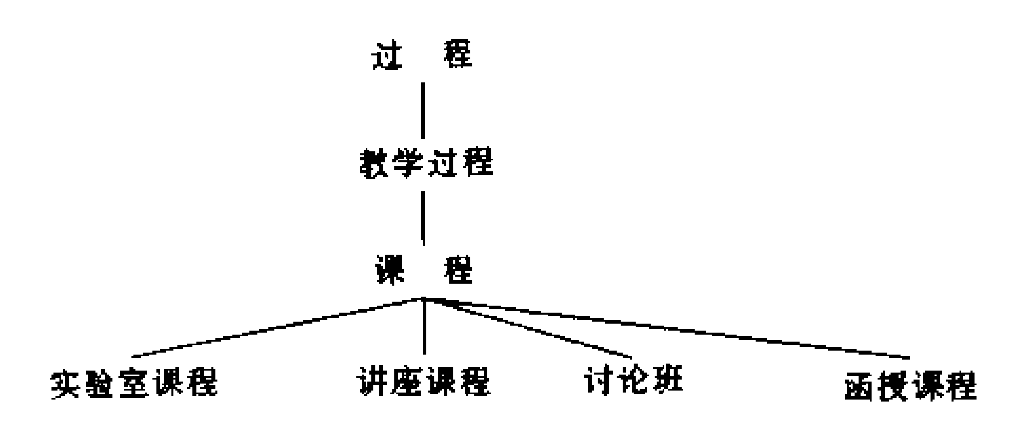


图 4.1 **课程**这一概念的分层次组织

课程，而后者又是一种类型的教学过程。这种概念的组织使得概念

具有非组织的规则集合所缺乏的在计算特性上的许多优点。对许多物理概念来说很重要的另一种类型的槽是**部分**，它建立了另一种类型的层次。例如，脚趾是脚的一部分，脚是腿的一部分，而腿又是身体的一部分。涉及部分关系的槽也可以转译成规则，例如**如果** *X 是脚趾*，**那么** *X 是脚的一部分*。但由槽和层次组织起来的概念具有计算上的优势，在下一节我们会讨论到。很显然，概念并不打算成为一个完备的心理表征理论。如果一门课已经满了，你还可以通过获得任课教师的签字认可而选上这门课，这条信息就不是课程概念的一部分，而是你对课程所了解到的一条规则。但概念所具有的计算性质上的优势，使其在模拟人类思维时能成为对规则的极有用处的补充。

计 算 力

组织化的信息封装到分层组织的概念中，这使得多种强有力的计算成为可能。大型的基于规则的系统面临的一个问题是如何挑选要使用的规则。一条相关的规则除非是能够从记忆中提取出来以供使用，否则对一个系统来说并无益处。在一个基于概念的系统中一个很有用的过程是**继承**（inheritance），通过利用**类别**槽所立起来的层次，有关概念的推导便可以很快地完成。讨论班课程会有教师吗？这个问题的答案很可能没有作为**讨论班**概念的一部分直接表示出来，但通过注意到讨论班是一种课程，而课程通常有教师，便可以很快得出讨论班课程有教师这一结论。这不是逻辑演绎，因为并非所有的课程都有教师。但预期讨论班课程有教师却是合理的，这一预期是由讨论班是一种课程继承而来的。

当你听到“课桌”一词时会想到什么？可能你会想到椅子、学习或台灯。并非所有的思维都如同基于逻辑和基于规则的系统那样是在做推理。人们可能会根据某些规则，如每一张课桌都配有一把椅子而将课桌与椅子联系起来。但这种联想可以根据课桌是

一类家具，而椅子是另一类家具这一情况而更富于因果性地产生。这种松散的联想在计算上被称为**传播激励**过程。系统的一个概念被激活，而激励可以沿着网络传播到通过类别和其它关系相联的其它概念上。激励传播就像电荷的传播，一种带电物体会使得与之相联的其它物体也带电。例如，如果某件事激活了你的**课桌**概念，激励可能会传播到你的**家具**概念（课桌是一类家具）和**抽屉**概念（抽屉是课桌的一部分）。这些概念的激活又可能导致一些其它相关概念被激活，如**桌子**（桌子是一类家具）和**木材**（抽屉通常是由木材做成的）。一些基于规则的系统也将传播激励吸收为一种激活规则的机制，以模拟人们怎样从记忆中提取规则（Anderson 1983；Thagard 1988）。

当用于应付新情境的时候，封装在概念中的信息最为有用。一门新课程开始后，学生们会很快将它置现入他们的概念系统，将它划分为一门讲座课程、一门小鸟课（特别容易的课）或别的什么。在这一过程中有两个关键步骤：**匹配**和**推导**。寻找最恰当的概念来对应一门课程，要求将各种相关课程的槽与关于这一门课所知的特定信息进行匹配。例如，如果一门课程仅有几名学生，这一信息就与**讨论班**概念的槽相吻合，讨论班的规模一般都比较小。如果教室里有 100 名学生，这便与**讨论班**概念的槽不一致了，但可能更适合**讲座课**概念。一旦一个概念与一个情境相匹配，学生们就可以通过转用该概念所产生的一整套预期来针对该情境进行推导。如果一门课程被归类为讨论班，学生就很可能会预计到会有很多次的课堂讨论。所以要理解到概念的计算角色，我们需要对一个处理系统考虑以下步骤：

1. 系统用激活的概念来表征一个情境。
2. 这些概念向其它潜在的相关概念传播激励。
3. 一些与当前情境相匹配的概念被选中。
4. 系统通过从被选中的概念进行继承来针对当前情境进行

推导。

问题求解

规划 当你首次遇到一个规划情境比如登记选修课程时，你可能需要使用普遍性规则进行搜索以得到一个解决方案。但第一次的成功会使你第二次的登记容易得多，因为你仅需按照相同的顺序再进行一次选课。你已经掌握了一个关于选课登记的脚本或者概念。这时的规划就不再是搜索或逻辑演绎，而是概念的应用。给定对你当前情境的表征及你所需要达到的目标，比如选上你所要选的课程，你便可以从记忆中提取能与当前情境及目标相匹配的选课登记的概念。这一脚本便会告诉你按适当的顺序去做什么：登记课程、交学费、如此等等。

然而，概念的应用仅仅是在你已拥有了与当前情境能很好匹配的有组织的信息封装后才有效。在大学学习了一年或更长的时间，学生们才有了一套对许多教学情境有效的程式，而一年级的学生可能就会将他们在中学掌握的程式错误地加以应用。选修认知科学导论课的同学有时会感到迷惑，因为他们预期此门课的内容会与他们所熟悉的哲学、心理学或计算机科学的课程相似。将一门跨学科课程纳入他们已有的概念里会有一定难度。对那些经常遭遇到的情境，脚本非常有用的，但如果当前的新情境与已有情境不相符时，则会对规划造成妨碍。想想这句格言：对仅有锤子的人来说，任何东西都像是钉子。所以你不应当把锤击的概念用到毫不相关的情境里。

决策 将概念应用到决策制定上的情况也是这样。在某些情况下，根据熟悉的脚本来制定决策不会造成麻烦，比如你总是从冷饮店选购某一种口味的冰淇淋。但人们经常会未经反思便使用熟悉的程式。例如，选择雇员的决策有时并不是基于对候选人是

否能恰当地符合雇人机构的要求的合理判断，而是因为某一特定的候选人与老板关于最佳雇员类型的概念相吻合。这一概念也许具有相应的槽，对理想雇员的智力和勤勉进行描述。但也可能包含了对于诸如种族和性别要求。因此尽管某些决策肯定是通过概念运用而作出的，而不要将此方法滥用到所有的决策制定上则幸莫大焉。概念运用是一种快捷而简便的决策制定方法，但不能照顾到对行动和目标的复杂的利害关系的考虑，而这些都是更为灵活的决策制定活动的一部分。

解释 与规划相似，解释有时也来自程式化的概念封装。特别是一些社会性的概念经常被用到解释里面，很可能超出了它们所适用的范围。为什么弗雷德通宵达旦地编写程序？因为他是一个电脑迷。为什么萨拉总是穿一身黑衣服？因为她是一位艺术家。为什么艾丽丝在一门她根本没用功学的课程得到了 A？因为这是一门小鸟（香肠，蛋糕）课。在所有这些例子中，解释都几乎是自动来自于将概念与似乎合适的情境相匹配的结果。

但概念还有更加灵活的解释性的用途。科学解释通常具有基于逻辑和基于规则的系统带给解释的那种**演绎**风格。例如，在物理学里，通常有一些普遍规律，如*力＝质量乘以加速度*，经过数学上的推演可以对行星的运动作出演绎性解释。但在很多领域内，比如进化生物学和社会科学里，规律是很难适用的。因而解释可以更好地概括为对程式的应用。程式包括一个目标——什么是需要解释的——和一种能够提供解释的模式。下面是采用达尔文的自然选择进化论来解释为什么一个物种具有某一特定性状的简化的解释程式（选自 Thagard 1994；也可参见 Kitcher 1981，Schank 1986）：

解释目标

为什么某一**物种**具有一种特别的**性状**？

解释模式

物种具有一套可变化的**性状**。

物种经受着环境的**压力**。

环境**压力**偏好该**物种**具有一种特别**性状**的成员。

所以该**物种**具有这种**性状**的成员比该**物种**缺乏这种**性状**的成员能更好地存活和繁殖。

所以渐渐地该**物种**大多数成员都具有了这一**性状**。

黑体字印刷的词汇是变量，可以填上多种不同的例子。例如，如果你要解释为什么某种细菌能够抵御抗菌素，使用这一模式时可以注意到能抵御抗菌素的性状是细菌物种的可变性状，抗生素导致了环境压力，所以能抵御抗生素的细菌会存活和繁殖得更好，直到整个物种都具有了抵抗性。

我们已经见到了好几个对认知科学来说是基础性的解释程式的例子。第一章的小结提供了一个基于表征和计算程序的一般化的解释程式，而第二章至第七章的小结则包括了各种特殊表征方式的解释程式。第二部分则讨论这种程式难以应用的有关心智和智能的各个方面。

学　习

我们知道对规则从何而来这个问题有三种回答：它们可以是天赋的，也可以来自于经验，或者由其它的规则形成。这三种回答同样也适用于概念：概念可以是天赋的，可以来自于具体的事例，或者由其它概念生成。不同的答案对应不同的概念。儿童学会新词汇和相应的新概念的速度大致是每天 10 个。

例如，**人脸**的概念由两只眼睛、一个鼻子和一张嘴组成。婴儿对这一概念的学习很可能就是通过重复接触人脸的示例的经验完成的。但有实验证据表明婴儿并不需要学习有关**人脸**的典型结

构，而是生来就预期人脸是一定的样式。与此相似，越来越多的证据表明一些基本的物理概念，如**物体**是天赋的，因为很小的婴儿就对物体应有怎样的行为显示出了很强的预期，例如物体消失在另一个物体的背后然后会重新出现。因而，虽说我们所有的概念甚至包括 VCD 和盒式磁带录像机这样的概念都是天赋的这一假设不尽合理，但一些基本的概念以及生成新概念的机制可能是我们与生俱来的心理装置的一部分。

一些概念是从具体事例中学来的，就像一些规则是通过归纳概括得来的一样。一部分概念必须从很多事例中吃力地得到，比如儿童学习区分狗和其它动物。不过，当你已掌握了很多概念，你就可以从小部分的事例中很快地获得新概念。假如你走进教室吃惊地发现只有几名学生，而且与讲座课不同，课上有很多的讨论，仅从这一个案例你便可获得关于**讨论班**的概念。当然，这个概念可能会在以后事例的基础上进行修正，正如规则通过重复的使用可以对内容和适用性进行细微的调整，概念可以随着经历事例的增多而得到修正。现已开发出许多通过事例形成概念的复杂计算模型（Michalski 等 1986）。联接主义的概念生成方法留待第八章讨论。

并非所有的概念都从具体事例而来，有时我们可以通过组合我们已有的概念来产生新概念。事例可以担任填补细节的角色，比如对**音乐电视**和**电子邮件**这类概念，但这些概念的内容的主要部分是由经组合而生成新概念的那些概念提供的。一部分概念的组合是直线式的，比如我们可以指出**宠物鱼**是作为宠物来养的鱼。但另一些概念的组合就复杂多了，例如**电脑迷**（computer geek）就不是既是电脑又是家伙（geek），而是指对计算机着迷成性的一类行为奇异的人。一些出乎意料的概念组合甚至涉及逆推推理的成份，需要提出假设来解释这一组合何以可能。例如，**盲人律师**这个概念不只是简单地由盲人和律师这两种属性组合而成，而需要增加新的属性如**勇气**来解释一位盲人何以能成为一名律师（Kun-

da, Miller 和 Claire 1990)。已有人研制了一些简单类型的概念组合的计算机模型 (Thagard 1988),但目前还没有更复杂的逆推类型的计算模型出现。

包含因果性信息的程式可用于进行一种逆推式推理。这里是得某种传染病比如伤风的脚本:

传染病

接触:你接触了某些病原菌(病毒或细菌)。

潜伏:该病菌繁殖。

症状:病菌致使你出现诸如流鼻涕等症状。

治疗:慢慢地,你体内的免疫系统杀灭了病菌。

假如你有像流鼻涕这类的症状,你可以填上症状槽,然后填上接触的槽逆推式地推断你一定是接触了某种病菌。

语　言

在口头和书面语言中,概念都是用词语来表达的。并非所有的概念都必须有词语来描绘它们,但在词语和许多概念之间确实存在一种很近的对应关系。在上一章里,我们通过语言学规则讨论过语法,然而,语言的知识显然不只是由规则构成的,我们需要掌握语词以便将它们插入到语法结构中去。字典里的词语表称为词典(Lexicon),所以表征在心智里的词语或概念集被称为**心理词典**。

乔治·米勒等人提出心理词典是分层组织的 (Miller 等 1990)。他及他的合作者们制成了一个巨大的电子词典 WordNet,有 60 000 多个英语词汇。像"狗"这样的名词是通过在表征力(图 4.1)一节中介绍过的类别和部分关系分层组织的。表达行动的动词如"注册"和"跑"具有另一种组织,根据它们行事的方式。例如,在某种意义上说跑是旅行的一种方式,而短跑又是跑

的一种方式。形容词如“容易”的组织结构又不一样。我们对语言的使用依赖于我们记住和使用相应于名词、动词和形容词的概念的能力。米勒（Miller 1991）探讨了心理词典的结构、词语是怎样形成的以及儿童的词汇量是怎样增长的。

语言的学习不只是学习语法规则；它还牵涉到发展一整套概念体系。乔姆斯基传统的语言学假定在语言和词典之间有一条明确的分界线，但这一区分受到了来自另一条路线的挑战，后者称为**认知语法**。兰格艾克（Langacker 1987）和拉科夫（Lakoff 1987）提出句法结构是与概念的本质和意义十分紧密地联系在一起的。

概念的意义是什么？它怎样对语句的意义发生作用？哲学家们对此曾深为苦恼，并且拿出了很多可能的答案。一方面，一个概念的意义又似乎来自于其它概念的意义，比如告诉儿童**短跑**的意义是一种快速的奔跑。另一方面，概念的意义又与对世界里的事物的观察联系在一起，比如儿童真实地看到别人进行短跑。一个概念的意义通常不会通过其它对概念的定义而得到，因为成功的精确定义是很少见的。同样概念的意义也不能由事例来完全说明，例如不能将**狗**的概念与一群狗等同起来。因此关于概念意义的理论必须包括对于概念怎样既与其它概念又与外部世界相联系的说明（参见第九章）。这两个方面对我们理解概念怎样成为我们使用语言能力的基础都是必不可少的。

心理学上的合理性

对于某种特定的心理表征来说，怎样去揭示它在心理学上的合理性呢？直接的方法是根据人们具有某种设定的心理表征的假设进行心理学实验而得到结果。间接的方法则是使用计算机对设定的心理表征进行模拟以解释涉及某些行为类型的心理学实验的结论。在第三章讨论的规则的心理学合理性的大多数证据属于第

二种、间接的方式。例如用基于规则的系统来模拟人对密码数字的问题求解过程。不同的是，多数有关概念的心理学实在性的证据来自于概念的心理学实验而不是计算机模拟。心理学家进行了大量的实验以检验概念的本质以及它们在分类中的角色。在这里我只能提到一小部分重要的实验。

在行为主义统治心理学的时候，几乎不提及概念或任何其它的心理表征。当对概念学习的研究自50年代开始时，普遍接受的传统观念是概念是经严格定义的（Bruner，Goodnow 和 Austin 1956）。不过，到了70年代实验证据的逐渐积累表明概念应当理解为典型化的条件而不是定义条件。定义条件提供的是严格的规则，比如我们说一个图形是三角形当且仅当它具有三条边。典型条件允许有例外，比如我们说狗一般有四条腿，尽管有些狗只有三条腿。原型（prototype）是指一组典型的条件，因而狗的原型就是："有四条腿，有毛，吠叫"，如此等等。按传统观点，将狗的概念用到某个特例如本杰身上就是检验狗的定义条件是否适用于本杰。而从原型的观点看，把狗的概念用到本杰身上是一种较为宽松的过程，看看狗的典型条件是否与本杰的特点相匹配。

心理学实验提示概念的应用更适合原型的观点而不是传统观点。波斯纳和基尔（Posner 和 Keele 1970）采用点的模式作为知觉范畴（Category），实验受试者先学习四种原型点阵模式的各四种变形模式，然后给出一组新的模式要求将其分类。对这些新模式，受试者发现那些与原型相匹配的最容易分类，而将那些与原型差别越大的模式进行分类，就要花费越多的时间并且产生越大的错误率。同样，人们确定"知更鸟更是鸟"比"鹅是鸟"更为快捷，因为知更鸟比鹅更接近鸟的原型。

里普斯、索本和史密斯（Rips，Shoben 和 Smith 1973）揭示出人们确实认为有些范畴比另一些更具典型性。例如，在北美香蕉与芒果相比是一种更典型的水果。当要求人们列出一个概念的示例时，他们会倾向于列举他们认为最典型的例子（Rosch 1973）。

如果要你说出一种鸟，你更有可能说出“麻雀”而不是“鸽子”。罗斯彻和摩维斯（Rosch 和 Mervis 1975）发现人们对一种鸟类其有多大典型性的判断对应于这种鸟所具有的那些最通常被赋予鸟的属性的程度，比如飞翔和筑巢。知更鸟和鸽子都是鸟，如果能给出**鸟**的定义，它们都会符合，但在认知层面上它们有较大区别，因为知更鸟比鸽子更接近于**鸟**的原型。

将概念视为原型有助于说明概念运用中的许多特点，包括人们所犯的错误。例如，当布瑞威尔和特雷因斯（Brewer 和 Treyens 1981）要求受试者回忆在他们等候过的某间大学办公室里有何物品时，受试者通常会错误地报告办公室里有书。书籍是**学校办公室**原型的一部分。心理学家还进行过实验以弄清概念组合的一些方面的问题（Smith 等 1988）。

对原型的发现与将概念视作类似于框架结构的计算观点是相吻合的。然而，有实验证据表明概念的结构不能完全由原型反映出来。巴尔萨诺（Barsalou 1983）和其它研究者提出概念比对一些典型性质的封装更为灵活、更依赖于语境（Context）。一些心理学家指出我们关于概念的知识与我们学习该概念时首次接触的事例有密切关系。这样，用一个概念就不是与原型进行匹配，而是将新的事例与旧的事例进行比较。概念的使用就与第五章将讨论的类比推理很相似了。

墨菲和梅丁（Murphy 和 Medin 1985）、基尔（Keil 1989）和其他研究者提出既不是典型特征的集合也不是具体事例抓住了概念的关键。对概念的使用有时除了像特征匹配之外还像是因果性解释，例如当我们看到有人全副衣装地跳入水池，我们会认定他喝醉了。全副衣冠地跳入水池并不是概念醉了的定义或典型特征，而是与有关不健全的判断力的理论相符合，而后者又是下述概念的一部分：喝醉了导致人做傻事。也许，我们应当指望概念也包含规则，比如**如果** X 醉了，**那么** X 具有的是不健全的判断力。作出一个人喝醉了的结论不只是与原型进行匹配，而是一种基于规

则的逆推式推理。昆达、米勒和克莱尔（Kunda，Miller 和 Claire 1990）在概念组合中发现了这种推理的证据。因此概念可能与规则、事例和典型特征都有密切的联系。

神经学上的合理性

在概念网络中的概念之间传播激励与神经元之间通过电化学脉冲的激励传播方式很相似，但目前对于概念如何在大脑中实现却知之甚少。大脑扫描技术正在用于研究语言的组织。波斯纳和拉彻勒（Posner 和 Raichle 1994）介绍了如何使用脑扫描技术对大脑响应诸如"锤子"等语词的过程进行监测研究。这些研究确定了涉及语词知觉和发声的大脑的不同区域。另一种研究心理词典的神经结构的方法是研究因中风造成大脑损伤的人在行为上的缺损。一位病人在说出无生命物体比如乐器的名称时有困难，却能较好地理解食物、鲜花和动物的名称。另一位病人中风后丢失了对水果和蔬菜命名的能力（Kosslyn 和 Koenig 1992）。人工神经网络（第七章）对于概念如何在大脑中存储和使用提出了一些看法。

实践上的可应用性

教育的职能之一是将初学者变为诸如物理学或其它科学分支领域内的专家。那么生手和专家之间的差别是什么呢？答案之一可能是专家掌握了更多的规则，但是教育学研究表明专家具有的是高度组织化的知识，这可以由概念或程式来加以描述（Bruer 1993）。例如初学物理的学生具有的有关斜面的程式只包含了一些表面的特征，如角度和长度。相反，专家的程式则将斜面的概念与相应的物理学定律联系起来了。内瑟塞（Nersessian 1989）和奇艾（Chi 1992）指出学生必须掌握一些抽象概念，如**场**和**热**，而学生却错误地把它们看成物质，这是科学教育的难点。

任何设计问题都涉及到可以由程式或框架所表征的概念。在建筑设计中，大梁的概念可以由框架来表达，有跨度、负荷、支撑强度和最大压力等槽（Allen 1992）。这个框架是一个涉及各种类型的梁的概念层次系统的一个部分，比如有钢制梁（I 型梁或盒型梁）、混泥上梁（增强型的或预应力的）以及木制梁。戴姆和列维特（Dym 和 Levitt 1991）介绍了一个名为 Sightplan 的专家系统，该系统用于对建筑现场的临时性装置的设置提供计算机支持。Sightplan 使用了框架来表达诸如**建筑现场**、**受力平面**以及受力平面的各个部分。

虽然不像基于规则的系统用得那么广泛，基于框架的系统在人工智能中仍具有相当多的应用。框架被应用到可能算得上当前最雄心勃勃的智能系统 CYC 中，这一系统试图编制大量的常识性知识，作为诸多领域的智能行为的基础（Lenat 和 Guha 1990）。CYC 拥有上万个框架表达许多日常概念和对象。由于最初的基于框架的表征方案有局限性，现在 CYC 也纳进了相当多由逻辑形式化表达的信息。纯粹的基于框架的专家系统是很少见的，但有些基于规则的系统也使用框架（Buchanan 和 Shortliffe 1984）。

小　结

概念，与口头和书面语言的语词有着部分的对应关系，是一种重要的心理表征方式。从计算和心理学上都有理由放弃概念具有严格定义这一传统观点，取而代之是将概念视为典型特征的集合。概念的使用则成为在概念与外部世界之间寻求一种近似的匹配。程式和脚本较之与词语相对应的概念具有更大的灵活性，它们在构成上有相似性，都是由特征集合构成，并可与新情境匹配，并应用于新情境。基于概念的系统的解释程式是：

解释目标

为什么人们会具有某种特定类型的智能行为？

解释模式

人们具有一整套概念，通过形成类别和部分层次以及其它联系方式的槽组织起来。

人们具有一套运用概念的程序，包括传播激励、匹配和继承。这些程序应用到概念上产生出行为。

概念也可以转译为规则。但概念对信息的组织与规则集不一样，支持不同的计算程序。

讨论题

1. 哪些概念是习得的？哪些概念是天生的？
2. 哪些概念可以定义？哪些概念具有典型特征？
3. 哪些概念不能与英语词汇相对应？哪些概念只能无意识地知道？
4. 概念可以归约于规则吗？规则可以归约为概念吗？
5. 基于概念的解释与基于规则的解释有何区别？
6. 你怎样表达心智这一概念？
7. 概念怎样与外部世界中的事物相联系？

进一步的推荐读物

对于心理学在概念上的研究工作的很好的评论，有 Smith 和 Medin 1981，Smith 1989 及 Medin 和 Ross 1992 的第十二章。Aitchison 1987 和 Miller 1991 提供了关于心理词典的导论性读物。对基于框架的 AI 系统的评论有 Maida 1990。

注　　释

分层组织的概念系统有时称为词义网。

Thagard 1992 年分析了在科学史上的主要科学革命过程中的概念转变。

第五章　类　　比

请想象一下如果每件事对你来说都是从头开始，你的生活将会怎样？如果每堂课都是你的第一堂课，每一天都是你的第一天，这会是什么样的后果？幸运的是，人们能够记住以往的经验，并从中进行学习。但这种学习并不需要形成在规则和概念中所见到的那种普遍性的知识,如果你是一名二年级或更高年级的大学生，你会记得你是怎样注册和选课的，这一经验或许是很有限的，使你不足以概括为一条普遍规则或概念，但你仍可以使用这一特定的经验来指导你在新学年的选择。如果你好不容易修完了一门对你来说是灾难性的课程，你会尽量避免再上一门同样类型的或有相似教师的课。而你若在一门课上获得了极大的成功，你又会再选上一门相似的课程。

类比思维是由采用你所熟悉的相似的情境去处理一个新的情境所构成的。人类对类比的使用可以追溯到最早的文学记载：荷马在《伊利亚特》中用到了类比，《圣经》中的寓言则在寓言故事与读者身处的情境之间提供类比。哲学家们很早就注意到了类比在推理中的重要性（例如，Mill 1974；Hesse 1966），但对类比进行心理学上和计算上的研究则是最近的事。艾万斯（Evans 1968）做出了第一个类比推理的计算模型，从那以后，众多的模型相继问世。今天，有好几个研究小组竞相研制使用类比的复杂模型，凯思·赫尔约克和我提出了一个关于人类类比运用的计算理论(Holyoak 和 Thagard 1995)。在后面的评价介绍中读者会看到，我们的观点与德雷·简特纳及其同事(Gentner 1989；Forbus，Gentner 和 Law，1995)所代表的流行观点既有相似也有不同。在目前的人工智能中，类比推理通常被称为**基于案例**的推理，并已开发

了许多饶有兴味的应用程序(Kolodner 1993)。道格拉斯·霍夫斯塔特和他的助手们(Hofstadter 1995; Mitchell 1993)则推出了关于创造性类比运用的新颖的模型。

表 征 力

类比能告诉我们除逻辑、规则或者概念之外更多的东西吗?就类比推理而言,我们需要能够表达两个情境,**目标**类比体用来表征将要推断的新情境,而**源**类比体则表征可采纳并应用于目标类比的旧情境。每一个类比体都是对情境的一个表征,而类比则是它们之间系统性的关系。对类比体的表征不仅要求对应用于个体的谓词比如"学生"予以注意,而且还要关注用以描述两个或多个个体之间关系的谓词,比如"教"。情境之间的类比具有相似的关系以及相似的特征。采用在第二章里引入的那种逻辑记符,我们可以像这样来表征 PHIL999 这门课的一些情况:

1. 任课教师(里帕索,PHIL999);即里帕索教授是 PHIL999 的任课教师。
2. 迟钝乏味(里帕索);即里帕索是个迟钝乏味的人。
3. 难(PHIL999);即这门课很难。
4. 登记选修(你,PHIL999);即你选修了这门课。
5. 得分(你,PHIL999,低);即你这门课的成绩得了低分。

此外,对你选修了一门教师迟钝乏味的困难课程而得低分最关键的是:

6. 因果(2&3,5);即迟钝乏味的教员和困难的课程导致了你的低分数。

在陈述6中，我们看到了一种涉及陈述句之间的因果关系的表征力，这对很多重要的类比来说往往是极为关键的。如果你在考虑选学PSYCH888，而这门课也有一名乏味的教员，且以难学著称，你可能会通过与PHIL999的类比而推断出你很有可能会得一个低分，从而避免选修这门课。在这里，PHIL999就是熟悉的源类比体，而PSYCH888是你在源类比体基础上所要推知的目标类比体。

在进行课程选择而不是避开某些课程时，你可以更积极地使用类比。如果有一门课你很喜欢，而且如果你能确认该课程的一些特征是导致你喜欢这门课的原因，那么你就可以寻找你还会喜欢的相似的课程。一个深思熟虑的类比使用者会撇开那些表面上的相似性，比如两门课程名称中字母的个数是一样的。但表面上的相似性又怎样与重要的相似性相区别呢？对一名学生来说，一天中的哪个时间上课可能无关紧要；而对另一名学生，上午是他学习效率最佳的时期，在选择与过去被证明是愉快的课程相似的新课程时，上课的时间就成了一个相关的因素。找出相关差别的关键之处便是鉴别出能导致与你选课这一目标有关的结果的因果关系。因此类比的表征需要包含如上面第6个陈述一样的对因果关系的表征。

通常，类比体可用于表征我们已经学过的各种表征方法的合集。类比体可以像概念而不同于逻辑和规则里的陈述句，将信息集束在一起而形成封装，但它可以包含描述某一个特定情境的信息，这又像简单陈述句，而不同于概念和规则。例如，（在本节开头）对PHIL999的表征就提供了对一门课程的信息封装，但封装内的信息只能应用于该课程，而不能普遍地使用。相反，在下面有关学习一节要讨论的类比程式，则包含了普遍性信息，这与规则和概念相似而不同于对源和目标类比体的表征。

类比有时可以表征为第六章讨论的那种视觉表象。图5.1提供了一个视觉类比。例如，当人们使用一幅熟悉的建筑物的心理

表象来猜测怎样在一栋不熟悉的建筑物里周游时，就利用了视觉类比。

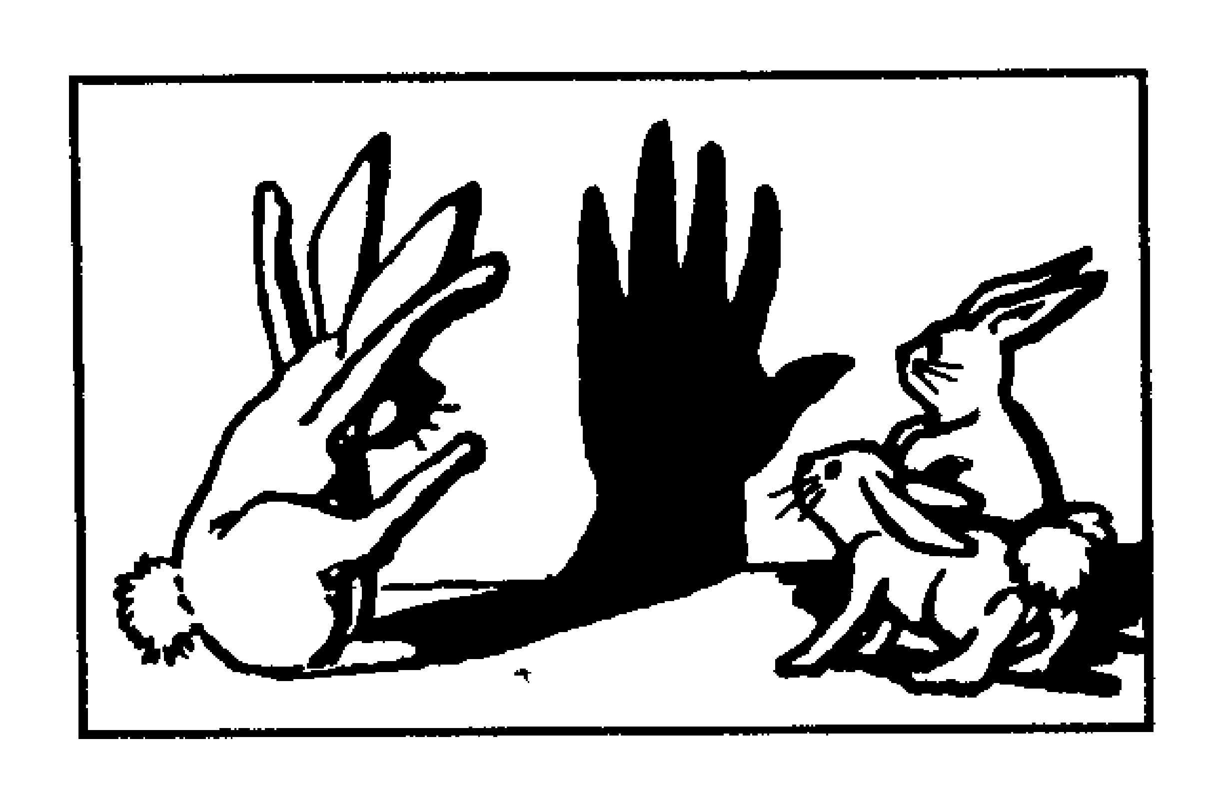


图 5.1　一个幽默性的视觉类征

经授权选自 Holoak 和 Thagard，1995，第 14 页。

计算力

当你在一个你已拥有相当多的专业知识的熟悉领域内求解问题时，你可以使用由规则和概念来刻画普遍性的知识，与此不同，类比推理则是当你对某一领域已具有一定的经验却缺乏普遍性知识时发挥作用，因此类比可以在概念和基于规则的知识难以奏效时显出计算效力。

通常，类比推理可以分四个阶段来进行：

1. 你面临一个目标问题需要解决。

2. 你记得一个已知解答的相似的源问题。

3. 你将源问题和目标问题进行比较，把它们的相关元素一一对应起来。

4. 你采纳源问题以对目标问题作出解答。

从计算上理解类比推理要求弄清楚记取（从内存中提取）、比较（将源类比体和目标类比体一一进行比较）和采纳三个阶段的程序。

从记忆中提取潜在相关的源类比体在计算上是非常困难的，在你的一生中会有多少经验？如果在过去的15年中你每天完成10个任务，那么在你的记忆中就存贮了54750个任务的解答。而对眼前的任务你要找到一个新的答案，你就要将新问题与数量非常之大的已存贮的解答进行比较。从这样巨大的存贮中如何选取可资利用的经验呢？假定你目前的任务为下一个学期或学年进行注册登记，你需要回忆起为某事进行注册登记的每一次经历吗？要记得你的每一次排队？你的每一次挫折？每一次你在雨中做的某件事情？

目前的进行类比体提取的计算模型在对什么是进行有效提取的因素以及什么是说明人们成功和失败地使用类比的标准等问题上存在分歧。在综合了许多研究者成果的基础上，赫尔约克和萨伽德（Holyoak 和 Thagard 1995）提出提取过程受到三个约束条件的制约：相似性、结构和目的。两个类比体如果涉及到相似的概念，那么在表面的层次上具有相似性。思考当前的注册事宜会让你想起以往进行注册的事例，以及其它在概念上与注册登记有关的行政事务。视觉类比体之间的相似性就不仅仅是概念上的了，还涉及它们在视觉上的外观。一辆小汽车会令人想起另一辆汽车是因为它们具有相似的外形或颜色。

然而，强有力的类比就不仅仅只涉及表面的相似性，还与深

层的结构关系有关，如果这一次的注册致使你错过了当天下午你所喜爱的电视节目，你也许会想起上一次你交纳学费也导致了你错过了电视节目。这两个情境之间的对应就不只是它们都牵涉到行政事务和错过电视节目，而是这种行政事务与导致你错过电视节目之间高度的相关性。为了完全满足结构性约束，两个类比体必须准确地排列为：

目　标	源
原因：注册（你）	原因：交纳学费（你）
错过（你，电视节目）	错过（你，电视节目）

在左边的目标类比体说的是你注册导致了你错过了电视节目，右边的源类比体说的是你去交纳学费而导致你错过了电视节目。虽然这两种情况一个是注册，另一个是交纳学费，它们都有同样的结构，因为所举出的“错过”和“原因”之间的关系是完全一样的。

提取的第三个约束条件是目的：你所要记起的案例能有助于你解决当前的问题。在人的记忆（以及计算机的数据库）里有大量的信息，所以要提出所有那些仅与当前任务相关的信息是一个心理学上和计算上都相当困难的问题，如果将类比的目的作为任务的约束之一，这一找寻源类比体并将其应用于目标类比体的任务，就会变得容易些。例如，如果你在注册和交纳学费使用类比的目的是揭示你所在大学行政机构工作效率的低下，那么这一目的会有助于你回忆起其它一些效率低下的情况。

赫尔约克和我指出这三个约束的并行操作才使得从巨量的相关信息中提取合适的类比体成为可能。其它研究类比的学者对此存有异议。福布斯、简特勒和劳（Forbus，Gentner 和 Law 1995）强调相似性在提取中的作用，而对结构和目的不重视。而另一方面，许多研究基于案例推理的学者则对提取与当前目的相

关的类比体感兴趣（Schank 1982；Kolodner 1993）。对建造专家系统来说，他们旨在研制一个对各个领域都有用的"通用索引词汇表"。关于人类的记忆是否以这种方式来编目只能留待心理学实验来检验了（见下文）。

一旦一个潜在的源类比体从记忆中被提取出来，它必须与目标问题进行对照以求找出能够对解答提供提示的对应之处来。如果两个类比体极为相似，对照就完全是细节问题了，正如你当前的注册与你以前经历过的一样。但是创造性的类比通常要有一个跳跃，如同在下面这个例子中所看到的（Dennett 1991，第 177 页）：

> 幼年海鞘在海底漂荡以寻找一块合适的岩石或珊瑚礁，然后就附着在上面并将它作为一生的居所。为完成这一任务，它要具有一套低级的神经系统。当它已找到了合适的地点并站稳了脚，它便不再需要它的大脑了，所以它便把大脑吃掉了！（这看起来就像是一位教授获得了终身教席）。

一位教授怎样像一只海鞘吃掉自己的大脑一样获得自己的终身教席呢？掌握这样的对比要求注意到一系列的对应：在海鞘和教授之间，在寻找岩石和获得终身教席之间，等等。认知科学家们在究竟是哪些约束在这样的对应中发挥作用的问题上存在分歧。简特勒（Gentner 1983，1989）强调这种对应是对结构关系的关注，但赫尔约克和我则认为表面的相似性和目的同样在类比对应中起作用。争论的双方都研制了计算模型用以检测相互竞争的理论主张。

如果源类比体与目标类比体能清晰地对应，那么将源类比体中的相关部分复制到目标类比体上即可求得解答。如果上一次你是通过修选一门晚上开的心理学课程来解决注册的问题，这次你也可以同样选择另一门晚上的心理学课。如果完全一样的解决不

可能，你可以对以往的解决方案稍加修改，比如选一门晚上的哲学课。对修改最复杂的分析是在基于案例的推理研究中。科罗德勒（Kolodner 1993）列举了十种修改以前解答的方法，从像用一门哲学课替换一门心理学课这样的简单置换，到较复杂的推衍，如在用某一种计算机语言编写程序时，系统地参照用另一种计算机语言写好的类似程序。

问题求解

规划 从上面的行文中不难看出类比怎样有助解决规划问题，比如选择好的课程。我们还可以将科学和数学课程中解答习题归入此类。课本的正文部分通常都包含展示怎样解题的例子，给定学生一些信息，然后要求求得答案。例如，给出一些有关一种化学物质的信息，要求你计算出诸如密度等这样的其他属性。类比并不是求解这类问题的唯一方式，但它对于解答每章背后的习题是很有帮助的．将书翻回去，将问题与课本正文中提供的已解答的问题联系起来。

类比对问题求解非常有用，但对一个新问题并不总能提供一个最佳的解决方案。选择一个与待求解的目标问题并无深层相关的相似性的类比体通常是很危险的。如果目标问题本质上是一个新问题，那么以往的任何解答都不会奏效，类比就只能带来误导。在制订军事计划时，使用过时类比体的将军们是会打败仗的。同样，虽然学生将新习题同以往做过的习题进行类比可以使解答大为简化，但如果新问题要求新的解答方法时，这一策略便会适得其反。在数学课上学到的技巧对于要求撰写论文的课程来说是极其有限的。

决策 决策是在不同的行动方案中进行选择，同样可以通过类比来进行。法律上的判决经常会援引以往的案例来作为先例：这

些案例作为源类比体，用来与当前的案例进行对照。历史学家记载了大量依据类比来进行政治决策的事件。例如，在1991年美国对是否要进攻伊拉克作为对伊拉克入侵科威特的报复进行论战时，支持和反对的双方都时常采用了历史上的类比。乔治·布什总统将伊拉克领导人萨达姆·侯赛因比作阿道夫·希特勒，暗示对伊拉克的进攻与二次大战时对德国的进攻一样，都是合法的。反对者则偏好另一个类比，将这一行动比作美国对越南的灾难性的介入。类比可以从提示过去的成功解决方案以及提醒决策者以往的灾难性的教训两个方面对决策提供帮助。然而，太常见的情况上，决策者往往只盯住一个以往的类比体而不去考虑有很多的源类比体可以提供不同的行动方案以供选择。

解释 类比对解决解释问题也是一种重要的资源，包括教师怎样将他们的知识传输给学生，以及在科学研究时怎样提出新的理论解释。不妨留心一下在你的下一堂课里，老师会采用什么样的类比。教师经常通过与学生已经知道的东西进行比较来帮助学生去理解他们不熟悉的东西。例如，对美国人去解释板球这种英国式体育运动时，我会将它与棒球进行对比，因为这两种运动都有击球棒（板）、球和跑垒。类比解释也有其限制，因为进行对比的事物之间既有相似也有差别，但对于使一个新手对一个新领域从了解到熟悉这一过程来说，类比无疑是十分重要的。在本章的后面部分关于教育的一节里会讨论如何在教学中有效地使用类比。

对认知科学本身而言，类比解释也是很丰富的。正如我们在前面已看到的，认知科学中的一个基本性的类比是介于心智与计算机之间的，我们试图把心智比作一台计算机去解释它是如何工作的。不过，这一类比很复杂，因为通过研究心智和大脑，我们对于计算可能是什么会产生出新的见解。关于计算的早期观点主要来自心理学角度，而第七章讨论的联接主义计算模型则受到对

大脑的新看法的影响。

学　　习

类比思维涉及三种类型的学习。最平庸的一种是仅仅在先前经验的基础上存下对案例的记忆。当你对一个问题已作出了解答，你可以把答案存到记忆里，这种存贮并不需引入类比，也不需要形成概念和规则等种种概括。但它是类比思维必要的前奏，构成了一种低水平层次上的学习。第二种学习是进行类比的直接结果，你采纳一个以前的案例来解求一个新问题，与我们讨论过的规则和概念相比，这是一种较为具体、特殊的学习，因为你所学到的只是怎样去解决一个特定的新问题。如果你是采用一个先前的解释问题来提出新的解释性假设，这种推理可以成为逆推式的。例如，如果一位朋友在晚会上迟到了，你可能会想起以前有一次某人晚会迟到是由于在来的路上车胎没气了，由此经类比而猜测你的朋友此次迟到可能是汽车出了麻烦。

第三种学习要引入一个概括性的元素。如果你使用了一个源类比体去解决一个目标问题，你可以对源和目标进行抽象而形成一个类比程式，来刻画它们二者之间的共同之处。例如，在你去年注册的基础上你解决了今年的注册登记问题，这样你便可以导出一个有关注册的抽象程式。类比性程式与上一章讨论过的程式（概念）非常相像，只是在普遍性的程度上要低一些，因为这只是从区区两个事例中所做的概括。在完成了两次注册登记之后，从这两个情境中你可以抽象出对注册的一个描述，其中包括涉及怎样选上你想修的课程的一些大致的规则。一个抽象的类比程式对未来的问题求解可能会非常有用，因为它包含了源和目标类比体共有的地方以及与问题解答相关的诸方面。在有关心理学合理性一节我们会看到形成类比程式可以增强解决问题的能力。

语　言

在生成和理解语言时，类比扮演了一个重要角色，因为它是运用隐喻（metaphor）的基础。当人们说库尔特·科拜因是90年代的吉米·享德里克，就不能从字面上理解为库尔特·科拜因就是吉米·享德里克，这只是指出了二者之间的某些系统性的相似性：都是摇滚歌星，吸食毒品且英年早逝。同样，说生活是战场带给人的是目标类比体（生活）和源类比体（战争）之间系统性的对比。而其它的隐喻，如生活是一场晚会，带来的又是完全不一样的另一种对比。信息高速公路并不是一条真实的高速公路，而是类比它能快捷高效地传输电子信息。

有些语言学理论家视隐喻不过是对语言的一种不正规的使用，只不过是没有依字而上的意义去使用语言，为什么不实话实说？相反，相当多的语言学家、哲学家和心理学家则把隐喻看成是语言的一个扩充性的、有价值的特点而不是一种例外或不正规的使用(Glucksberg 和 Keysar 1990；Lakoff 和 Johnson 1980)。所有的隐喻都具有类比映射所特有的那种系统性对比的认知机制，虽说隐喻可以比类比更多地使用其它形象化的手段从而产生出更宽广的联想范围。不论是言者说出隐喻还是听者理解隐喻都要求以类比为基础的知觉。如果我对你说里帕索教授是一只海鞘，你应当能理解我并不是说他是一种具有囊状身体的海洋动物，而是说他的心理历程与海鞘的生活经历有着某种相似性。

心理学上的合理性

关于人们怎样使用类比已有许多的心理学实验。在这里我只谈一些类比在问题求解、学习和语言使用方面的例子。

如何解决框盒5·1中所提的问题呢？大多人觉得很难有办法让医生使用射线消除肿瘤而不毁坏健康的身体组织：吉克和赫尔

约克（Gick 和 Holyoak 1980）发现只有 10%的大学生能够找出好的解决办法。

框盒 5.1　肿瘤问题（选自 Gick 和 Holyoak 1980）

假定你是一名医生，一位病人的胃里长了一个良性肿瘤。不可能对这位病人施以手术，但如果肿瘤不除去病人就会死去。有一种射线可以用来消除肿瘤。如果射线能以足够高的强度突然照射到肿瘤上，肿瘤就能被除去。但以这种强度照射的话，射线到达肿瘤前穿过的健康的组织也会被损坏。低强度的射线对健康组织无害，但同样也对肿瘤无效。采用什么方法能够用射线既消除肿瘤，同时又不损伤健康组织？

在听完框盒 5.2 里的城堡的故事后，75%的大学生能够找出一个好的解答。初看起来，城堡的故事与肿瘤问题毫无关系。但很多人能够使用城堡故事里把军队分散然后再集中攻击城堡这一解决方案对肿瘤问题提供解答：不必使用单股的高强度射线，医生可以从不同的方向用多股低强度的射线去消除肿瘤。

这个例子说明的是一种简单的类比学习，即通过采纳一个旧的问题去解决一个新问题。借助同样问题，吉克和赫尔约克（Gick 和 Holyoak 1983）研究了学生怎样从多个事例中学习类比程式。除了城堡故事之外，还教给部分学生一个有关消防队员的故事，灭火员通过使用多个小的水龙头来扑灭着火的油井。大火被会聚集中的水扑灭，正如城堡被会聚集中的部队占领。听了两个故事的学生在教师指导下对两者之间的相似性进行了反思，他们在解答肿瘤问题时比只听了一个故事的学生更能记起使用一种会聚集中式的解决方案。所以学习类比程式有助于更有效地解决问题。

涉及语言的心理学实验是用以研究隐喻的使用。格鲁克斯伯格和凯伊萨（Glucksberg 和 Keysar 1990）的实验结果显示当人们被要求按字面意思去理解时，他们还是去找寻隐喻的意义。在一项研究中，大学生们被要求判断一些句子如“有些人的书桌是垃圾场”在字面上是否正确。在对一个字面上错误但还有隐喻性解

释的句子（如上例）正确地回答“不是”时，反应的速度要比回答一个字面上意义错误而缺乏隐喻性解释的句子来得慢，如“某些书桌是马路”。相似的发现也可以从可同时从字面上和隐喻式的方式进行解释的语句中找到。凯伊萨（Keysar 1990）给学生呈现诸如“我儿子是个婴儿”这样的语句，要求学生判断这个句子在字面意义上或者隐喻意义上是对还是错。学生被要求尽可能快地按键来表示该语句在字面上是正确的。如果语句在字面意义上是错的，而它在隐喻意义上也是错的，学生们的判断会更快；如果该语句在字面意义上是正确的，而在隐喻意义上也是正确的，学生的判断也会更快。隐喻性解释看来像是伴随字面意义处理的一个强制性过程，而不是在字面意义处理之后的一个可选择的过程。

框盒 5.2　城堡故事（选自 Gick 和 Holyoak 1980）

一个小国处在一位独裁者的铁腕统治下。独裁者在一个坚固的城堡里统治这个国家。城堡位于国家的中央，周围是农场和村庄。从城堡出发有多条道路向外辐射就像车轮的辐条一样。驻守边疆的一位了不起的将军集合一支大部队并立誓要占领城堡，把国家从独裁者手中解放出来。将军知道如果他的全部人马能够立即全力进攻城堡就能够攻下城堡。他的部队集结在一条通往城堡的道路上，准备出发去攻城。这时一名间谍给将军带来了一条令人不安的消息：残暴的独裁者在每一条道路上都埋设了地雷。地雷埋设后只有小股人马能够从道路上安全通过，这是因为独裁者需要从城堡里调动军队和民工出入。但是大部队通过就会触发地雷，这不仅会破坏道路，使得军队无法通过，而且独裁者还要血洗很多村庄作为报复，对城堡实施大规模的直接进攻看来是不可能了。

然而，将军并没有被吓倒。他把他的部队分成很多的小组，每一小组分别到通往城堡的不同道路上。当所有的小组都准备就绪后他发出了信号，每个小组沿着不同的道路向城堡进发。所有的小组都安全通过了雷区，接着部队全力进攻城堡，就这样，将军占领了城堡，推翻了独裁者。

神经学上的合理性

迄今为止，似乎还没有关于类比思维的神经学基础的任何证据。

实践上的可应用性

正如我们在计算力一节所看到的，类比对于解释可以提供实质性的帮助。所以类比在教学方面具有很大的利用价值。好的教师经常会将学生不熟悉的东西与他们所熟悉的东西进行比较，来帮助他们进行理解。不过，在教学中使用类比也有许多潜在的危险。要避开那些陷阱需要注意到学生们所知道的是什么，以及类比是怎样被使用和误用的。

这里是为教育工作者更好地使用类比所提供的一些简明的建议（更详细的讨论和辨析，见 Holyoak 和 Thagard 1995）：

1. 使用熟悉的类比源。在解释科学问题或其它复杂的、不熟悉的东西时，使用同样不熟悉的东西来作类比，显然是不足取的。如果孩子们并不了解太阳系的结构，你用与太阳系的类比来解释原子的结构是不可能的。

2. 尽量使对应变得清晰。对一个好的类比，学生们应该能够自己指出源类比体与目标类比体之间的对应关系，但提供一些指导有助于发现二者之间的对应。例如，说到认知科学时指明心智那些方面与计算机的哪些方面相对应是十分重要的。

3. 使用深层的、系统性的类比。不同于表面特征的对比，强有力类比使用可以为学生要解决的问题提供清晰对应的系统性因果关系。

4. 介绍误匹配。任何类比或隐喻都在某些方面是不全面的或

存在误导。某些教育学家由此认为类比有太多的误导而不宜用于教学，但问题的解决不是拒绝使用类比，而是要指出在什么地方发生了偏差。没有人会期待在信息高速公路上会有白色的斑马线。

5. 使用多重类比。当一个类比出现问题时，加进另一个类比可以对上一个类比的不完整之处提供补充说明。

6. 进行类比治疗。找出学生使用的是什么样的类比并在必要时予以校正。

这些建议不仅是对教学方面的的类比运用有用，而且还能应用到对类比的其它使用，包括问题求解和决策的制定。

类比是创造性设计的一个丰富的源泉。乔治·德·梅斯特尔（Georges de Mestral）在观察到芒刺是如何粘在他的狗身上后发明了尼龙拉带，亚历山大·格拉汉姆·贝尔（Alexander Graham Bell）从人耳的构造受到启发发明了电话。工业设计家们经常使用反向式工程的技术，将竞争对手的产品分解开并仿制出类似的产品。

类比和隐喻同样对计算机设计和人机交互方案的构思有所帮助。PC 机的 Windows 软件（类比式地）仿制了 Macintosh 的界面，而后者又借用了一个关于办公桌的类比：屏幕像一张办公桌，用户在上面摆放各种公文和文件夹。帐目处理软件使用一组格式来进行帐目计算，这类似于老式的记帐簿。文字处理器软件在某些方面就像其取而代之的打字机。

产品设计的效用有时也会被用户未经思索的类比所妨害。一位妇女对一家计算机公司报怨道尽管她已用脚把鼠标器踩在地板上，计算机仍不启动。她大概是把计算机当成缝纫机了。一个男子报怨说他把一页文件举在屏幕前面计算机却没能发出传真，他显然是把计算机想成了复印机。设计人员需要考虑到有助于消费者使用的积极的类比，但也要留心用户自己冒出来的引人误入歧途的类比。有时消费者可能需要进行类比治疗。

科洛德勒（Kolodner 1993）介绍了数十种基于案例的推理系统。虽然它们在特定的提取和映射机制上各有千秋，但基本上都是在已有解答的基础上运用类比推理以解决新问题。部分基于案例的推理系统目前已投入商业应用(Allen 1994)。例如，Lockheed公司使用了一个名为Clavier的基于案例的推理系统来对如何在一个称为自动转炉的大型高压传送炉中安排飞机的零部件提供建议（Hinkle和Toomey 1994)。使用的案例（源类比体）是对以前安置于高压炉的负载的记录。高压炉专家无法将他们的专业知识用规则表达出来，但把他们的经验做成一个能存贮、提取和采纳案例的系统，却能满足Lockheed公司的日常需求。

小　　结

类比在人类思维中担当了一个重要角色，在各种不同领域如问题求解、决策制定、解释和语言交流中都发挥了作用。计算模型能模拟人们怎样提取和映射源类比体以便把它们应用于目标情境。类比的解释程式如下：

解释目标：

为什么人们会具有某种特定类型的智能行为?

解释模式：

人们具有对情境的语言和视觉表征，可以用作案例或类比体。

人们具有在这些类比体上进行提取，映射和采纳的操作过程。

这些类比过程，运用于类比体的表征上，产生出行为。

相似性、结构和目标这些约束有助于克服怎样找寻以往的经

验以及使用经验来解决新问题所带来的困难。并不是所有的思维都是类比式的，而且使用不恰当的类比会对思维造成妨碍，但类比在教学和设计等应用领域仍是非常有效的。

讨 论 题

1. 类比体（案例）与规则和概念有何区别？

2. 类比式问题求解在什么时候最有用？

3. 类比式思维有哪些主要阶段？在每个阶段最主要的制约因素是什么？

4. 借助类比进行思维潜在的主要危险是什么？

5. 类比如何对创造性提供帮助？创造性的其他来源是什么？

进一步的推荐读物

Hall 1989 回顾了当时人工智能在类比方面所做的工作。Holyoak 和 Thagard 1995 则是一个较为偏重心理学的综述。Gentner 1989 回顾了她在类比研究上的工作。Kolodner 1993 是关于基于案例的推理的一本很好的人工智能方面的书籍，也可参见 Riesbeck 和 Schank 1989 及 Schank，Kass 和 Riesbeck 1994。Hofstadter 1995 提供了一个有关他的研究小组在创造性类比方面的研究工作的引人入胜的回顾。

备　　注

最近的一些类比模型已使用了第七章介绍的联接主义（神经网络）的方法。见 Thagard 等 1990 和 Holyoak 和 Barnden 1994。

第六章　表　　象

在你居住的住宅楼或者公寓楼的正面有多少个窗户？你怎样来回答这个问题？如果你从未数过，现在你必须想办法来数一下，也许你会把楼房正面的所有房间制成一张表，然后在口头上数一下有多少个窗户，但很多人回答这类问题是通过形成一幅心理图像而且做目计。同样，请回想一下你是怎样从家里来到学校的，虽然你可以对此有一个完整的语言记忆（“沿主大街到十字路口再向右转”），很多人则会构造一系列沿途的道路、建筑物和其它路标的心理表象来记住这样的路径。

很多的哲学家，从亚里士多德到笛卡尔和洛克，都认为类似图画的表象是人类思维的一个本质部分。在19世纪后期现代心理学诞生不久，像威廉·冯特这样的研究者就研究了人们怎样用表象来进行思维，有些人甚至宣称没有表象便没有思维。20世纪初行为主义的兴起致使谈论心理表象和其它内部表征在科学上成为不光彩的事。但认知心理学在60年代的勃兴又使得心理表象再次成为科学研究的合适对象，像派依维奥（Paivio 1971）及舍帕德和梅兹勒（Shepard 和 Metzler 1971）等研究者开始使用心理表象来进行心理学实验。由于得到了许多实验的支持，视觉表象的计算模型开始出现（Kosslyn 和 Shwartz 1977；Funt 1980）。也有部分认知科学家对人类思维涉及与语言表征不同的图像表征持怀疑态度（Pylyshyn 1984），但大量计算、心理学和神经学上的研究表明心智的思维与借助语词一样借助了图像。

虽然认知科学家们对表象的兴趣主要集中在视觉表征上，我们也不应忽视表象与非视觉表征也有关。胡椒比萨饼的滋味是什么？如果你曾经尝过，你也许会形成一幅有关滋味和气味的心理

表象，并用它来判断其他东西，比方说“潜艇”三明治是否有与胡椒比萨饼相同的味道。长胡子的下巴摸上去像不像砂纸？为了回答这个问题，你可能会对这二者生成一幅触觉表象并进行比较。你是怎样把棒球击到对方半场，单手运（篮）球，或者清洗一面镜子的呢？如果你对这些体育或身体活动有经验，你便能形成一幅与这些活动有关的身体感觉的运动表象。最后，我们还可以有情绪表象：当你得知你被大学录取的时候你是什么感觉？朋友们的感觉一样吗？第九章会讨论情绪和意识。本章的余下部分则集中讨论视觉表象，这是目前研究得最多的一种。

视　　觉

对于具有正常视力的人来说，看东西似乎是一个自动而又轻而易举的过程，你向房间里看了一眼，飞快地看出了里面有人、家具等。然而，当你想让计算机来做这一切时，视觉的复杂性就立刻突现出来了。将一台摄影机对准房间然后把图像分解成像素——组成电视屏幕上一幅图像的那些点——存贮起来，这并不困难。但是从成千上万个像素中抽取信息却非常困难，因为摄影机获取的图像可能是相当含混不清、模棱两可的。如果有一个人坐在一张椅子上，那么通过摄影机获得的像素就只能展现椅子的一部分，所以计算机必须能推断出那是一张椅子，虽然它无法看到与一张标准的椅子完全匹配的全貌。房间里某些部分可以比其它部分要明亮一些。墙上的长方形物体可能是一幅图画，也可能是一面镜子，照出房间的其它部分。在过去的几十年间，计算机视觉取得了可喜的进步，使得机器人在简化的环境中能够辨认和操作特定的对象，但与人类的视觉能力比起来，机器人的视觉仍显得相当粗糙。

看看图 6.1 的那幅图画，如果你在这个正方体的顶部和底部之间来回转移注意力，你能使得这个立方体前后变换，将第一个面看成正面然后又将另一个平面看成正面。这种情况是如何发生

的？这个画面的反射光进入你的眼睛照到你的视网膜上，视网膜由数百万个光感受细胞组成。但是在你的大脑将这幅画面解释为一个立方体之前有着许多的处理加工过程，大脑必须检测边缘，从背景中区分出线段。在图 6.1 中边缘检测不困难，但如果在亮度、灰度和黑白对比度上有细微的变化，这就不是件容易的事了。而且，大脑接收的并不是像摄像机所生成的那种单一的图像，而是从双眼获取信息．双眼看物体时有细微的差别，而这一差别使你能够觉察物体的距离并看到三维物体。

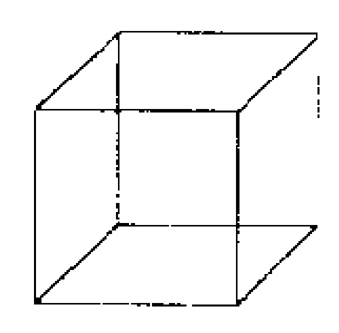


图 6.1　勒克尔立方体

最上端的边缘既可以看成是在立方体的前面，也可以看成是在立方体的背面，请试着将注意力集中到不同的边缘使得立方体前后变换。

大脑通过将边缘检测和不同的视角、颜色及其它信息结合在一起而得到一个对物体的一致性的解释，这些物体远比图 6.1 中的立方体复杂得多。关于视觉信息加工可进一步参考马尔（Marr 1982）和科斯林（Kosslyn 1994）。所有这些加工过程的结果是一幅视觉表象。这样的表象并不完全取决于显现到我们眼前的那个物体，因为我们可以将表象存贮在记忆中，提取它们，并对它们进行各种操作，以满足各种心理任务的要求。

表征力

为什么人们平时会说一幅图画相当于千言万语？图画可以用

语言来加以描述，我们来看一看图 6.2，我们可以说一个人坐在另一个人后面，从打开他的头颅往里看。使用足够多的语句，我们可以提供一个更为详尽的描述。但图画式表征却有诸多的优势，语言描述可以包含这样的信息：坐在椅子上的人紧挨着坐在沙发上、头颅被打开的人，这样我们可以从字面上推断出椅子与沙发是挨着的。而用图形来表征，这样的推导就毫无必要了：我们就能直接看到椅子与沙发是挨着的。图 6.2 是一个我们用眼睛去看的外

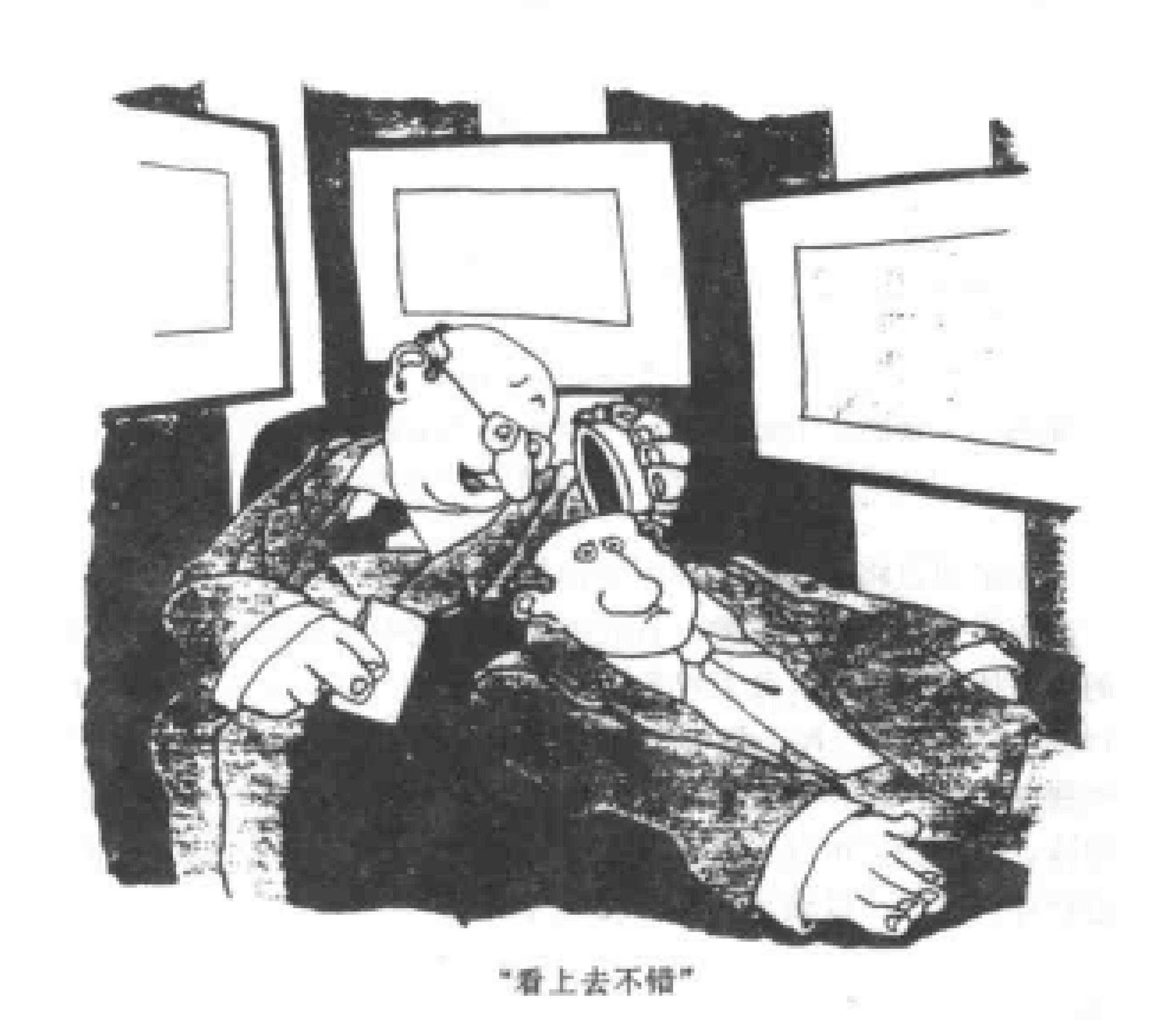


图 6.2　选自《纽约人》上的卡通画

由盖汉·威尔逊创作。© 1994，《纽约人》杂志公司。经授权引用。

在的表征，但如果你把这幅画面合上一会儿，你依然能够形成这幅画面的一个心理表象，并能回答有关它的一些问题。这两人的年龄大概是多少？他们之中有人秃顶吗？有人戴领带了吗？

图画和视觉心理表象对于表达事物看上去是怎样的以及它们的空间安排提供了强有力的方式，但并非所有的信息都能自然地用图画来表征。抽象语句如“正义是公平”就无法在视觉上进行表征，而概括性语句“所有的恐龙都已灭绝”用图画来表达就十分笨拙。同样，因果性语句“吸烟导致癌症”和“如果你感冒了，那么你会咳嗽”就无法用画面来直接表达。因此视觉表象对我们在前面章节里看到的那些语言表征方式来说是一种补充而不是替代。

在前面的章节里，我们假定了表征本质上是语言式的：对规则、概念和类比的讨论都是用语言形式来表述的，但这些结构也可以采用视觉图像的形式，一条规则可以有**如果**（图画1）**那么**（图画2）这种结构，就像电影一样用画面2跟在画面1之后。概念也可以是图像式的，比如我可以将狗的原型不表征为特征集合而用具有这些特征的一张狗的图片来表达。同样，源和目标类比体可以具有如图5.1所示的兔子和影子那样的视觉表征，因此，除了语言式的规则、概念和类比，还存在着视觉式的规则、概念和类比。

心理表象的结构是什么？科斯林（Kosslyn 1980）及格拉斯哥和帕帕戴斯（Glasgow 和 Papadiass 1992）提出心智使用的是类似阵列的结构来执行视觉任务。例如，我们可以使用图6.3所示的阵列来代表欧洲。最近，科斯林（Kosslyn 1994）提出人的大脑使用各种神经网络来表征空间信息（参见下面神经学合理性一节）。

计算力

许多可以由表象来完成的思维活动也可以由语词来完成，但

对有些任务语言思维就来得吃力多了，视觉思维对那些依赖视觉外观或空间关系的问题显得十分有用。视觉表征，无论是心理的还是外在的，比语言表征更能支持多种不同类型的计算程序：

				瑞典	
苏格兰 威尔士 英格兰			丹麦		
		荷兰	德国	德国	
		比利时			
	法国	法国		克罗地亚	塞尔维亚
葡萄牙	西班牙				希腊

图 6.3 用一个阵列表征的欧洲地图

经授权引自 Glasgow 和 Papadias1992，第 373 页。

1. 检阅 想象在一只盘子的左边有一把刀子，其右面有一把叉子，那么刀子是在叉子的左面还是右面呢？答案可以根据“左”、“右”关系的逻辑属性从语言上推导出来。但可以更直接了当地通过看一下形成的表象而看到刀子在叉子的左边，这个程序也可以检查两个表征对它们进行比较。

2. 查阅 你把鞋放在家里的什么地方？为了记起这个，你可以对你的房间做一个心理扫描来找到它们可能存放的地方。

3. 缩放 蛙有尾巴吗？一些人在回答这个问题时是通过生成一幅蛙的心理表象然后将它的后半部放大以便看得更仔细些，正如你可以把一幅图片拿近些看一样。

4. **旋转** 把一个大写的字母"E"放平会是什么样子?回答这个问题的一种办法是在心里面把字母旋转至平躺着。

5. **变形** 请按照芬克、平克和法拉(Finke, Pinker 和 Farah 1989)的指令来做:想象字母"B"并将它左旋 90 度,把一个与旋转后的"B"的直边长度相等的三角形头朝下放在"B"的下边,然后把水平线段移走。很多人把最后结果的图形想象成一个心形或是有双峰的圆锥形冰激凌。我们能以有力的方式改变和组合视觉表征,包括翻转、并置以及旋转。

这 5 个操作使得表象能以同前面章节介绍过的语言类表征大不一样的方式完成多种问题求解。为了回答是否你所有的鞋都具有相同的鞋带孔数目,你可以提取一幅你鞋柜的表象,对它进行扫描找到你的鞋,将其放大以检查你的鞋,接着将鞋的表象作一变形将鞋面部分并置在一块儿,以便比较鞋带孔的数目。而如果你仅有一双鞋,同时你知道同一双鞋中的两只具有相同的孔数,这可能会更容易地推演出答案而毋须求助于心理表象。

问题求解

规划 假设你有很多杂事要办:到杂货店购物,寄一个包裹,然后去送要干洗的衣物。通过前面的章节你能够以语言方式作出一个如何合理有效地完成这些任务的规划。一套如果-那么规则可以指导你去杂货店、邮局和干洗铺,或者是根据你以前的经验用类比或程式指导你完成这些任务。同样,你也可以通过视觉方式形成一个规划,想象你驱车进入杂货店的停车场,然后出来奔邮局,最后到干洗店门前停车,这样一个视觉规划可能要用到记录了你所要去的地方的空间关系的一幅心理地图。并不是人人都用到这样的心理地图,有些人使用语言编码的路标会干得更好,但对很多人来说,通过使用视觉表象来确定他们所在之处以及如何

到达所要去的地方，会更方便一些。

使用视觉表征进行规划包含的步骤与第二章介绍的基于规则的问题求解是相似的，只不过是通过视觉方式来进行的。首先必须对初始状态和目标状态形成视觉表征，然后构造一条从起点到目标的视觉路径。在解决构造问题时视觉变换是很有用的，诸如怎样在河的两岸之间建一座桥，甚至更现实一些的科学问题，解题者通常使用图表作为心理表象的外在的临时性辅助工具，例如，在几何学里画出图形和角度的草图对解题是非常有用的。学生在解答一些科学课程的习题时也经常使用图表来把握复杂对象，如弹簧、分子和染色体。

决策 关于表象对决策制订的贡献目前还很少有研究。但假定你要决定是穿你的蓝色夹克还是穿棕色的夹克，你可以想象它们与你要穿的其它衣服搭配起来效果怎么样，这样关于穿什么的决策就是一个视觉表征比较的结果。同样，如果在餐馆里决定要点什么菜，你的决策就部分取决于你对不同菜肴风味的想象了。在第九章我们会看到，情绪表象对决策制定也很重要。

解释 对于作出解释来说，视觉推理可以发挥很大的作用。据说伟大的发明家尼古拉·特斯纳 (Nikola Tesla) 在诊断非常复杂的机械故障时，只需要在脑海里形成一幅机器的心理表象，将它运转一下就能知道毛病出在哪儿。视觉解释在心理学或人工智能上都还未得到充分的研究，但我们有理由相信无论是在科学思考和日常思维中它都是十分普遍的。看一看世界地图上非洲和南美洲这两块大陆，然后把这两块大陆拼到一起，使巴西的凸出部分与西部非洲能镶合起来。本世纪初，这两块大陆外形的契合给了阿尔费雷德·魏格勒以启发，使他想到这两块大陆可能曾经是连在一块儿的。于是他提出了大陆飘移假设，来解释它们是如何分离的。这些假设可以完全用语言来陈述，但用视觉效果来表达非

洲与南美洲的契合是最好不过的，而且可以由两块大陆的视觉连合来加以解释。这种连合倒回去便是对很久以前发生的两块大陆的空间分离的猜想。如同规划一样，视觉解释不是语言解释的替代物，而是一种很有价值的补充。

学　　习

运动员们经常使用表象进行训练以提高他们的运动技能，有实验证据表明如果与实战训练相结合，心理表象的练习可以促进运动能力的提高。准备表演跳水或击打棒球的选手可以先想象一下如何完美地完成任务，这既要用到视觉表象也需要运动表象。将这些任务先在脑海里过一遍可以使你在实战时完成得更好。

表象对于概括也很有用，比方人们可以借助某一范畴中成员的图片例如大象的图片来产生一般化的关于大象的心理图像。对一头大象的视觉表征忽略关于某些特殊大象（例如象背上坐着骑手）的偶然性信息，而留下了普遍性信息（如：灰色、皮肤有皱褶等），表象式的概括学习还未受到实验研究或计算研究的重视。

逆推式学习也可以是视觉式的，如果你在车门上发现了一条很长的划痕，你会对此作出各种语言解释。但你也可以有一部心理电影：某人在停车场上从你的车边驶过，他的车门开着划到了你的车上。通过构成一系列画面来表明你车门上的划痕从何而来，你对另一辆车划过你的车所作的逆推式推理就是借于视觉表象作出的。其它的画面也是可能的，比如有人推着手推购物车从旁边经过，或是有人用钥匙划的。谢雷（Shelley，印刷中）介绍了考古学家在他们对古代物品作出解释时怎样使用视觉逆推。

语　　言

语言在本质上是言语式的，那么表象又怎样与对语言的使用有关呢？在上一章我们看到语言不仅仅只是句法和简单的语义学的事．对语言的使用经常是隐喻式的，正如拉科夫和约翰逊

(Lakoff 和 Johnson 1980) 已指出的，许多的隐喻有视觉上的来源：他今天**起**床了，她正处于她事业的**顶峰**。拉科夫 (Lakoff 1994) 提出相当多的理解涉及表象程式，后者是具有视觉元素的普遍性概念。例如，在对范畴理解的背后是对容器的视觉理解：一个对象可以在一个范畴**里**也可在一个范畴**外**，比如对范畴 “**狗**”，也可以被**放进**一个范畴里或从一个范畴里**移出**。隐喻还可以将不止一种的感觉表征拴在一起，如过分张扬的服装 (loud clothes)。

朗盖克尔 (Langacker 1987) 提出了**认知语法**，将隐喻和表象作为心理世界的中心，包括对语言的处理。他认为感知表象在概念结构里担当了实质性的角色；例如，“喇叭” 的意义可能部分地与它所产生的声音联系在一起，虽然这种语言学路线是有争议的，但它提示我们语言怎样依赖于视觉和其它表象。

心理学上的合理性

有很多的心理学实验支持视觉表象是人类思维的一部分这一主张。科泊和舍帕德 (Cooper 和 Shepard 1973) 测量了学生在判断一个旋转的字母是正常的还是镜像式的所用的时间。图 6.4 显示了字母 “R” 的各种样式，第一个 R 是正常的，第二个则是镜像反射式的。通过心理旋转可以看出第三和第四个 R 可以分别是镜像反射的和正常的。如果字母旋转的角度不是太大，如 5 和 6，确定它们是正常的还是镜像反射式的所用的时间比 3 和 4 的情况要少。

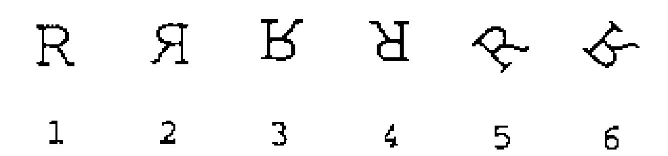


图 6.4 心理旋转：确定一个字母是正常的还是镜像式的所用的时间量与旋转它们以找到答案的角度成正比

除了旋转实验，扫描实验也证实了心理表象假说：人们在表象上扫描的距离越大，所花的时间就越多（Kosslyn 1980）。先建立一幅你的国家的心理表象，在东西海岸或边界各指定一个城市，在中部再指定一个城市，例如，美国人可以在地图上标出旧金山、芝加哥和纽约。如果你采用视觉表象，那么你从西部城市扫描到东部城市比你从西部城市扫描到中部城市所用的时间要长。

芬克、法拉和平克（Finke，Farah 和 Pinker 1989）进行的实验表明人们能够给不完整的或经变形的表象赋以新颖的解释。除了上面介绍的将旋转的 B 变成心型的实验外，他们还对学生给出如下的指令：想象字母“Y”，在它的下部放上的一个小圆圈，在它的半腰处加上一条水平线段。然后将图形旋转 180 度，在这一系列指令结束后，大多数人报告说他们看到的是一个杆状人形。这一系列变换如图 6.5 所示，人们获得正确答案的频率之高提示我们人们使用视觉表征进行操作。

虽然大多数心理学家认为上面提到的这些实验表明人类使用视觉表象是可信的，但也有一部分人持有怀疑态度，他们坚持所有思维的基础是同一种语言式表征，而表象的经验都只不过是错觉。旋转、扫描和其它变形都可以由词语表的非表象式计算程序来模拟。然而，在最近 10 年间，不断积累的神经学方面的证据为表象假说提供了进一步的支持。

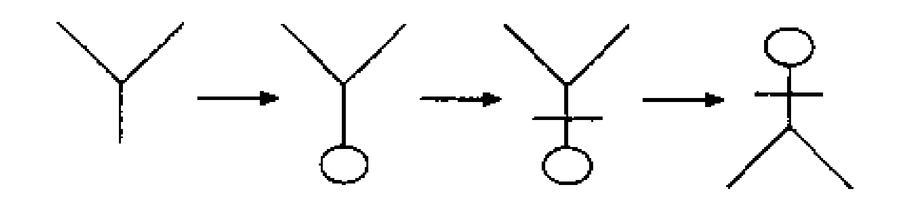


图 6.5　产生杆状人形所要求的变换系列

神经学上的合理性

科斯林（Kosslyn 1994）集中讨论了大脑用于视知觉的部分同样涉及视觉心理表象的两种类型的证据。首先，因大脑损伤而造成知觉能力缺损的病人有时也具有同样的表象缺损。例如，有些病人在知觉时不能看到视野的一边，他们在运用表象时也同样不能看到这一边。枕叶的受损会导致视觉表象的缺损。其次，对大脑活动的测量表明当人们使用视觉心理表象来执行任务时，大脑用于视知觉的区域也同时被激活。表象所依赖的大脑皮层区域在空间组织上是与视网膜的结构相对应的，视网膜是向大脑输送脉冲的神经网络。大脑与视网膜直接相连的区域在空间组织的结构上与视网膜相似。由于这些区域保留了外界对象呈现给视网膜的某些空间结构，它们在运用表象时的活动提示表象涉及类似图像的表征，而不仅仅是语言式的描述。

科斯林介绍了大脑使用能并行满足多重约束的计算机制来对心理表象进行的加工过程。第七章会讨论这样的过程如何由人工神经网络来完成。

实践上的可应用性

如果心理表象对于问题求解有用的话，在教学中更有效地使用表象无疑会使教育获益匪浅。拉金和西蒙（Larkin 和 Simon 1987）介绍了图表有助于有效的问题求解的情况。不过，大多数有关表象的心理学研究还集中在人们如何使用表象而未涉及如何在教学中更好地利用表象。

依照视觉表象来增强记忆有许多策略。为了记住某件重要的事情，可以将它与一幅生动的画面联系起来。例如，为了记牢本书讨论的六种类型的心理表征，你可以将它们分别与一幅不同动

物的画面联系起来。

许多建筑师、工程师和产品设计师的设计使用了诸如框图和蓝图等视觉表征，心理表象很可能是这些设计师创造性心理活动的一部分，但还没有多少心理学证据或计算方面的研究涉及表象在设计中的角色。芬克、瓦德和史密斯（Finke，Ward 和 Smith 1992）讨论了表象对创造性发明的作用。

虽然人工智能对表象和基于图表的系统的兴趣在不断增加，但基于表象的专家系统还很少。福布斯、尼尔森和法汀斯（Forbus，Nielson 和 Faltings 1991）介绍了一个对物理器件作定性空间推理的系统，格拉斯哥、福梯尔和艾伦（Glasgow，Fortier 和 Allen 1993）使用了一个基于阵列的系统来鉴别晶体和分子的结构。

小　　结

视觉和其它类型的表象在人类的思维中担当了一个重要的角色，与传统的语言式描述相比，图画式表征可以以一种更为可用的方式抓住视觉和空间信息。适合于视觉表征的计算程序包括检阅、查找、缩放、旋转和变形，这些操作对于适合用图画表征领域中的规则和解释的生成极为有用。视觉表征的解释程式是：

解释目标

为什么人们具有某种特定类型的智能行为？

解释模式：

人们具有对情境的视觉表象。

人们具有在这些表象上进行操作的诸如扫描和旋转的加工过程。

这些构造和操作表象的过程产生出智能行为。

表象对于学习以及某些具有表象根源的语言隐喻的使用有所帮助，心理学实验提示诸如扫描和旋转等视觉加工过程借助于表象，而最近的神经生理学研究的结果证实在运用心理表象进行推理与知觉之间有紧密的物理联系。

讨论题

1. 内省是研究心理表征和程序的可靠方式吗？为什么单凭内省不足以揭示心理表象的重要性？

2. 你具有感知表象吗？什么时候你用到它？

3. 哪些计算在使用表象式表征时易于实现？

4. 视觉表象对哪些类型的问题求解有用？在什么时候它们又成为了障碍？

5. 对心理表象的批评者怎么解释支持心理表象的心理学和神经学实验？

进一步的推荐读物

Kosslyn 1994 对最新的神经学和心理学成果进行了全面的回顾。Finke 1989 总结了很多表象方面的心理学工作。Glasgow 1993 评价了从计算角度关于表象的争论，与 AI 方面的批评者进行了探讨。Tye 1991 提供的是一个哲学上考察。Langacker 1987 探讨了表象与语言学之间的关联。Marr 1982 是有关人类与计算机视觉的经典性文献。

备　　注

表象的计算模型相对缺乏的重要原因之一是当前的编程工具更适合于语言式表征。除 Glasgow 和 Papadias 1992 提出的阵列表征外，图形表征对刻画视觉表征的某些方面也有用（Wong，Lu 和 Rioux 1989）。

第七章　联　　接

在 19 世纪即将结束时，桑狄亚哥·罗蒙·Y·卡贾尔发现大脑是由分离的细胞组成的。这些神经元通过称为突触的特殊节点的接触相互间传递信号，图 7.1 显示了一幅简化的由突触联接起来的神经元的画面。人的大脑有大约 1000 亿个神经元，很多神经元与上千个其它神经元联接，形成了神经网络。

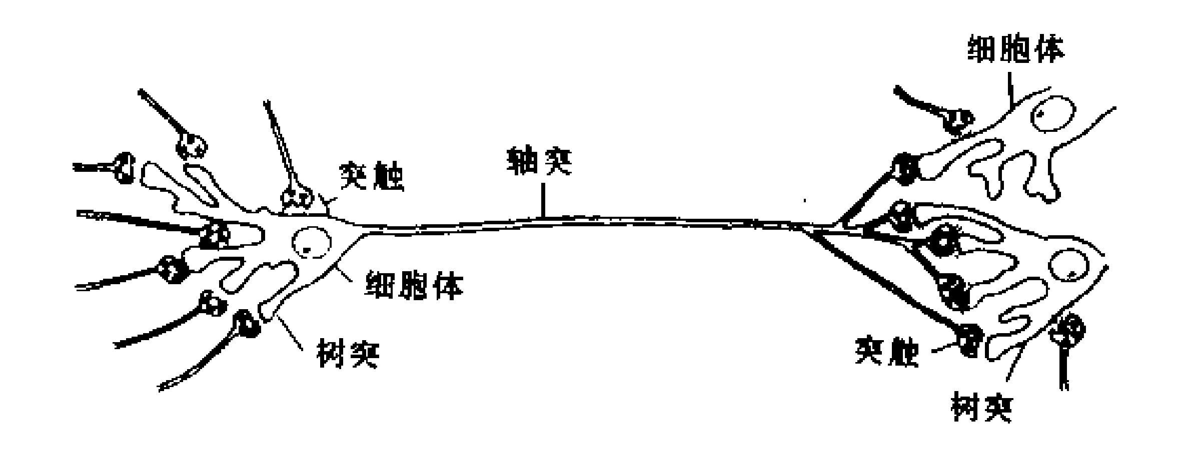


图 7.1　由突触联接的神经元

电信号从树突流入再从轴突流出

经授权选编自 Rumelhart 和 McClelland 1986，第 2 卷，第 237 页。

在 50 年代和 60 年代思维的计算模型研究的早期岁月，不少研究者的兴趣集中在模拟神经网络怎样完成思维的任务。这项工作到 70 年代随着人工智能和心理学领域的研究者的注意力几乎全部转移到基于规则和基于概念的表征上，从而走向了衰落。然而，在 80 年代，受大脑神经结构的启发的计算模拟研究又得到了

戏剧性的复兴（如 Hinton 和 Anderson 1981；Rumelhart 和 McClelland 1986）。这一研究纲领通常被称为联接主义，因为它强调简单的类神经元结构之间联接的重要性，此外，有时也被称为神经网络或并行分布式处理（PDP）。目前已经研制了相当多的联接主义的心智和大脑模型，但我将集中讨论两种类型，第一种涉及**局部式**表征，其中的类神经元结构被赋予了特定的概念或命题的可确定的解释。第二类涉及在网络中的**分布式**表征，对概念或命题的表征是以更复杂的方式进行的，即把意义分布到类神经元的结构的综合体上。

局部式表征和分布式表征都可以用以**实现并行约束满足**，许多的认知任务都可以从计算上理解为同时满足多个约束条件的加工过程。作为约束满足问题的一个简单例子，不妨考虑一下学校的教务人员所面临的安排全校一学年的新课表的问题。他们面临的有些约束是硬性的：他们不能够在同一个时间给同一个教师安排两门不同的课程；相反，许多约束却是软性的，包括任课教师和学生对他们的课程安排在何时何地的优先性。考虑到可供使用的教室以及教师与学生和优先性所引发的各种各样的约束条件，排出一个合适的课程表是一件令人头痛的问题，通常未能得到优化的解决。教务人员一般是依据前一学期的课程表，再根据需要进行修正以满足新的要求，但如果将所有的约束条件同时加以考虑，约束满足问题可以以一种更为普遍的方式加以解决。

并行约束满足的明晰的模型最初是为计算机视觉研制的，马尔和波吉约（Marr 和 Poggio 1976）提出了用于立体视觉的程序，他们称之为“协作”算法。两只眼睛产生的是对外部世界的略微不同的表象：大脑如何匹配这两幅表象并构造出一幅一致性的合成表象呢？马尔和波吉约注意到匹配是受到好几个约束条件支配的，涉及到一幅表象中的点怎样与另一幅进行对应。这种建构一幅一致性的表象就成了对在两幅表象之间进行匹配的约束条件的满足。为了从计算上完成这一任务，马尔和波吉约提出使用一个

并行的、相互联接的处理器网络，其中用联接机制来表征约束条件。

随后相似的网络被费德曼（Feldman 1981）用于模拟记忆中的视觉表征，麦克兰和拉姆哈特（McClelland 和 Rumelhart 1981）也用作字母知觉的模型。回过去看一看图 6.1 中的勒克尔立方体，并行约束满足提供了一种机制来解决隐含在勒克尔立方体中的二义性问题，两种全局性解释均可以由一套对图形元素的更基本的解释来定义。例如，在一种解释中图形的左上角是立方体正面的左上角，而在另一种解释中同一点则解释为背面的左上角。进而言之，可能的局部性解释是高度相互关联的，按照所认定的典型立方体的结构关系，要么相互支持，要么相互竞争。

人们对勒克尔立方体的解释可以由一个简单的联接主义网络来模拟，用单元来代表立方体的角，单元之间的联接代表解释之间的相容与不相容。在这一网络中，并行约束满足聚敛于两种可能解释的一种或另一种，激活的一部分单元合起来表征一种一致性的解释，未激活的则表示另一种解释。过去十年的大量研究已表明并行约束满足可应用到很多种高层的认知活动中，不仅仅只针对视知觉。

表征力

联接主义网络构成的表征非常简单，因为它仅由单元和联接组成，单元类似于神经元，具有激励等级，这与神经元激活时用以向其它神经元发送信号的频率相对应。在局部式联接主义网络中，单元具有可说明的解释，比如特定的概念或命题，一个单元的激励值可以解释为对一个概念的可用性或一个命题的真值的判断。联接可以是单向的，激励从一个单元流向另一个；也可以是对称的，激励可以在两个单元之间来回流动。联接要么是兴奋型的，一个单元增进另一个单元的激励值；要么是抑制型的，一个

单元压制另一个的激励值。图 7.2 给出了一个局部式网络的简单例子，它可以对一名学生的情况作出推断。你认识艾莉斯并且知道她喜爱编程，所以你会想到她可能是个电脑迷，而且性格内向不善交际。而另一方面，你又了解到她喜欢参加舞会，这说明她性格开朗。为了形成对她的一个一致性的印象，你必须判定她到底是性格内向还是性格开朗。图 7.2 中的网络用单元来代表每一种特点；并且使用单向兴奋型联接使激励从已观察到的行为流向推断出的性格。在性格内向和性格开朗之间还有一个对称的抑制型联接，反映出二者的水火不容之势。下面将要介绍的分布式网络中的单元则没有这种可以指明的解释。

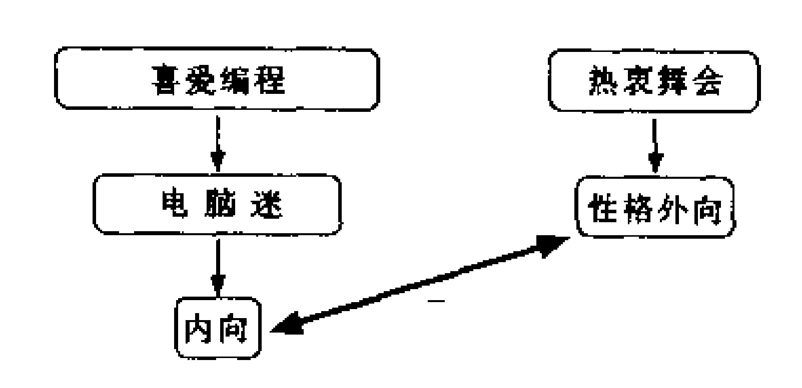


图 7.2　带有兴奋型联接（细线）和抑制型联接（带有减号的粗线）的简单局部式网络

哪些兴奋型联接有可能是对称的？

为了理解分布式表征的实质，我们来看一看由科斯林和科伊尼格（Kosslyn 和 Koenig 1992，第 20 页）提出的一个视觉类比。图 7.3 显示的是一个由章鱼组成的网络，其任务是当池底有鱼出现时向水面上的海鸥报信。当底层的章鱼发现了鱼便收缩它们的触角向中间层的章鱼发出信号，而中间层的章鱼以同样的方式向最上层发出信号，最上层的章鱼则伸出触角通知海鸥。这属于一种**前向式**网络，信息自下而上通过网络。底层的章鱼可以看成是

一个输入层，最上层看成是输出层，但中间层的章鱼又怎么解释呢？有关池底有多少条鱼的信息不是由任何一个特殊的章鱼来编码的，而是分布在整个章鱼的网络之上。同样，在图 7.4 中，这

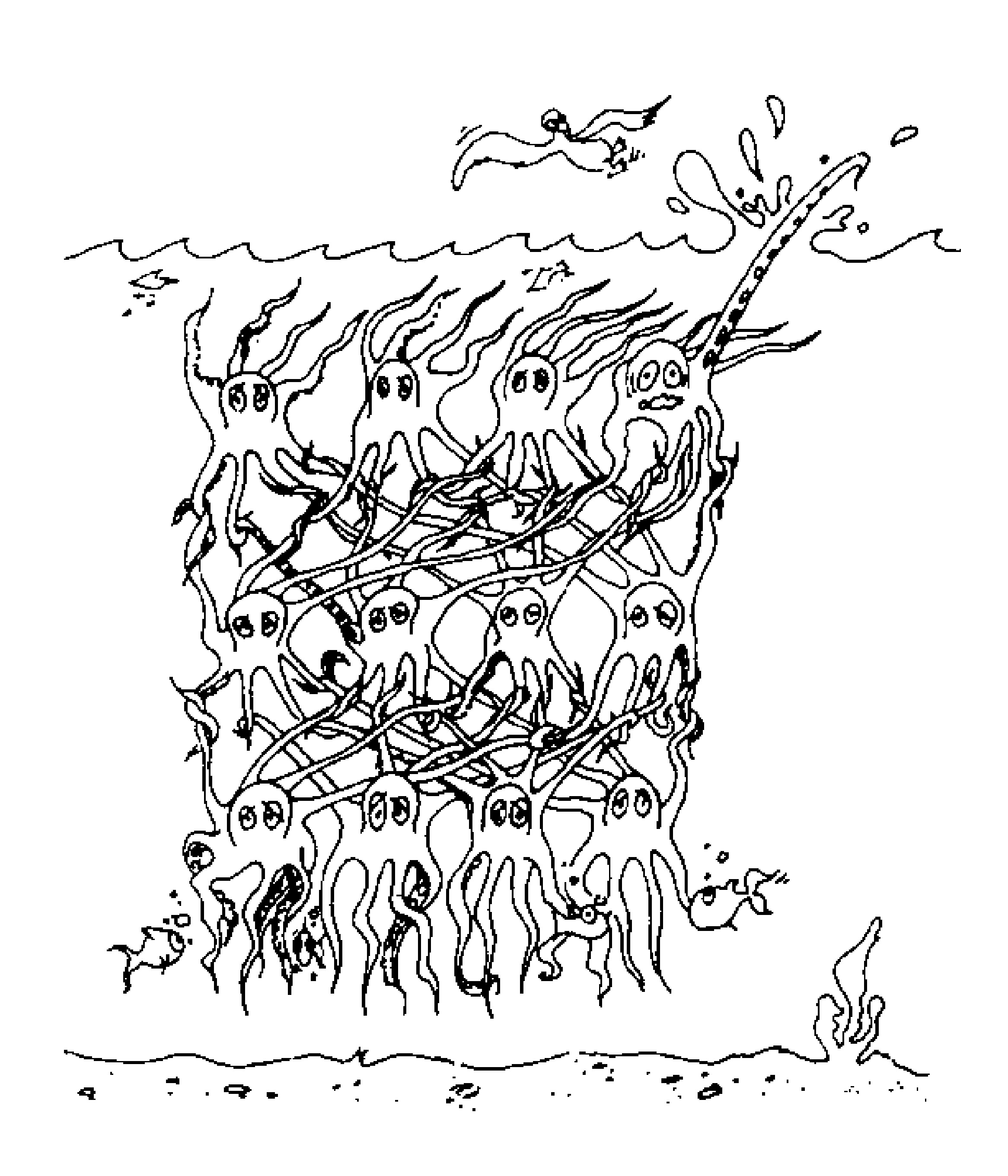


图 7.3　分布式处理网络的一个视觉类比

经授权，引自 Kosslyn 和 Koenig 1992，第 20 页

个前向式神经网络的中间层的隐单元（既非输入单元也非输出单元）没有任何初始性解释。它们是通过调节它们与其它单元的联接权重来获得解释的，这是通过下面将要讨论的学习过程来实现的。

概念（第四章）可以看作是网络上的分布式表征。一个经训练后对刺激作出准确反应的网络可以获得对应于该刺激的概念。例如，如果一个网络的输入单元用来检测动物的特征，而输出单元用以确定动物的种类，比如狗、猫，等等，那么该网络就能获取关于狗或猫的概念。这一概念不是由某个特定结点来代表，而是由当给出一组典型特征作为输入时出现的一个典型的单元激励模式来表征的。在一个分布式网络上作为结点激励模式的概念表征同在第四章给出的概念表征很不一样，但是在对概念是原型而不是一组必要充分条件的认识上是相同的。

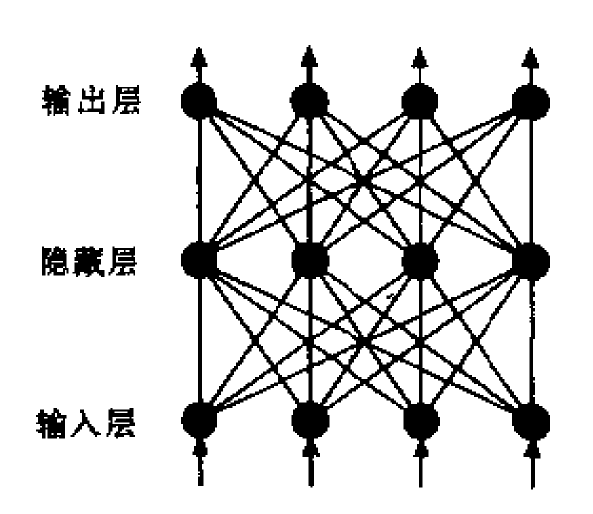


图 7.4　一个具有三层单元的前向式计算机模型

经授权，引自 Kosslyn 和 Koenig 1992，第 21 页

单元之间的联接对于表征简单的联想是够用的，比如电脑迷性格内向，而性格内向的人不开朗。但它们缺乏刻画较为复杂的

规则的表征力，比如某人行为举止与一个电脑迷很相似，那么他也是一个电脑迷。用第二章的逻辑符号表示：

$$(x)\{(\exists\ y)[\text{电脑迷}(y)\ \&\ \text{相似}(x,y)] \rightarrow \text{电脑迷}(x)\}$$

在字面上是：对任何的 x，如果存在 y 是一个电脑迷且 x 与 y 相似，那么 x 是电脑迷。像“相似”这样的关系和更复杂一些的逻辑关系很难用联接主义网络来表征，尽管研究者们也在进行很有益的尝试以增加胜过图 7.2 所示的简单局部网络的表征力。一种有希望的技术是使用同步方式来联接表征联想元素的单元。代表 x 与 y 相似的一个单元或一片单元可以用与代表 x 喜爱电脑同样的时序模式来激活 (Shastri 和 Ajjanagadde 1993)。然而，即便是表征力有限的网络也可以完成强有力的计算。

神经网络所提供的强有力的感觉表征使我们能比使用通常的语言更好地表征味觉和嗅觉(Churchland 1995)。舌头有 4 种感受器，甜、酸、咸和苦。假设一个系统各有一个单元来对应这 4 种感觉器，每一个单元有 10 种不同层次的激励水平，那么对应于每一种不同的激励模式，这个系统就可以辨别 $10^4=10000$ 种不同的味觉。

计 算 力

问题求解

为了对并行约束满足是如何进行的有一个印象，我们可以想象拔河比赛。同一般的拔河比赛略有不同，我们这里的拔河比赛不但可以拉，还可以推，而且有多个队同时参与，不只限于两个队。不仅如此，各个队之间还是相互连锁的，即一个人可以在多个队中。我们把这种游戏叫作“拉-推”比赛，每位选手要帮助同伴站立住，同时要把对手推倒。每位选手与其他选手之间要么是用绳子，要么是用棍子联接起来，用绳子互相拉而用棍子互相推。每个人以各种方式与其他人联接起来，由绳子联起来的两个人组

成一个队里的一对：如果一个人倒下了，那么他（她）便会拉着同伴往下倒；如果一个人站起来了，他（她）又会去拉起他（她）的同伴。同一队中的联接对相互间的作用力是一样的。相反，由棍子联接起来的两个人是对手队中的成员：如果一个人站起来了，那么他（她）就要努力把对手推倒；如果一个人倒下了，另一个就会更容易地站起来。相互联接起来的选手就构成了一个约束满足网络的内部结构。

现在假设再加上另一个作为“拉-推”比赛的指挥，指挥与一部分选手用绳子联接，可以拉；而与其他选手用棍子相联，可以推。但没人能影响指挥，他（她）总是站着的，指挥作为对这个约束网的外部影响，能够对一些选手施加拉力或推力，由他们再去影响其他人。渐渐地，在一系列的推拉之后，一些选手最后站立住而其他人则倒下了。选手们逐渐形成了分类，要么站着要么倒在地上：互相支持而站着的那些队获胜，而倒下的那些队输掉了比赛。

这种“拉-推”游戏类似于并行约束满足的实现，选手代表了元素而绳子和棍子则代表约束。一旦元素和约束都已确定，在一个并行网络中实现这一模型就非常容易了。首先，元素由单元（选手）来表征；其次，正的内部约束（绳子）由兴奋型联接来代表：如果两个元素之间的关系是正约束，那么代表元素的单元之间应当由兴奋型链路加以联接；第三，负的内部约束（棍子）由抑制型联接来代表：如果两个元素之间的关系是负约束，那么代表这两个元素的单元应由抑制型链路来联接。第四，外部约束可以通过将代表满足外部约束的元素之单元与一个对应于“拉-推”比赛指挥的特殊单元相联接来加以刻画。特殊单元影响的单元要么通过兴奋型链路与之形成正性联接，要么通过抑制型链路形成负联接。选手们相互推拉直至一些人最终站定而获胜的这一过程类似于单元间激励传播直至网络进入一个稳定状态的过程。一些单元最终处于激活状态，恰如一些选手最终站定着，而其他单元

处于非激活状态，如同其他选手被推倒在地。这一最终结果取决于单元（选手）之间的相互联接。

约束得以并行地满足是通过在所有单元之间反复传递激励，直到若干循环后所有单元达到稳定的激励水平。这一过程称为**释放**，类似于物体逐渐形成一个稳定的形状或温度的物理过程。获得稳定性被称为**安定**（settling）。释放网络意味着在单元互联的基础上调节所有单元的激励，直至所有单元具有或高或低的激励值。一个与处于激活状态的单元通过兴奋型链路相联的单元会从中获得激励，而通过抑制型链路与一个激活单元相联而会使自己的激动值降低。

规划 在竞争性规则中进行选择可以很自然地通过并行约束满足来理解，而构造计划则通过规则或类比来理解则显得更自然一些。你毕业的计划可以由一组规则来表述，包括如何安排课程的次序以便你获得所规定的学分，联接主义网络可以实现一些简单的基于规则的规划。陀雷茨基和欣顿（Touretzky 和 Hinton 1988）建造了一个使用分布式表征的基于规则的系统，它将与规则的如果部分进行匹配的过程作为一种并行约束的满足。不过该系统只能对简单的谓词而不能对关系型的条件子句进行匹配。内尔逊、萨伽德和哈迪（Nelson，Thagard 和 Hardy 1994）使用局部式表征实现了作为并行约束满足的规则匹配和类比运用。该系统能够实现规划构建，比如莎士比亚戏剧的朱丽叶如何设计与罗密欧幽会的方案。因此联接主义系统可以间接地与解决规划问题有关。

决策 我们可以通过并行约束满足来理解决策的制定过程（Thagard 和 Millgram 1995；也可参见 Mannes 和 Kintsch 1991）。决策中的元素是各种行动和目标，正内部约束来自于促进关系：如果一个行动促进了目标的实现，那么该行动与目标是一致的。负

内部约束来自不相容关系，两个行动或目标无法同时实现或满足，比如你无法在同一时间上两门课。决策制定的外部约束来源于目标的优先性：有些目标具有更强的吸引力，提供正约束。一旦对一个特定的决策问题确定了元素和约束，我们就可以使用在“拉-推”游戏中介绍过的基本方法去构成一个约束网络。

假定你面临的是在毕业后选择工作这样一个困难的问题。你可能的选择包括升入研究生院或进入一家大公司得到一个初级职位。你首先面临的约束是你无法两项都选，而且不同的选择适合你不同的目标。立即就业可能可以解决你当前的经济困难，但不一定能给你提供适合你兴趣的长期职业。而且，在你的领域中可能还有很多东西是你想要学习的。而另一方面，你可能已经厌倦了上学读书。图 7.5 是一个反映了这一决策问题的一些要素的简单网络。单元代表各种选择和目标，加号和减号表明反映基本约束的兴奋型和抑制型联接。如果将一个单元置为高激励值，可以解释为它所代表的行动或目标可以接受，而非激励值则表示拒绝。代表研究生院的单元比代表参加工作的单元具有更强的兴奋型联接，因而得到更多的激励，而后者由于与前者之间有抑制型联接面被减弱激励值。

类比在决策制定中也同样有用，因为一个以往的由 A 引起 B 的案例，可以使人认识到 A 有助于 B，使得使用类比推论本身也依赖于并行约束满足。本书第五章介绍了赫尔约克和萨伽德(Holyoak 和 Thagard 1995)的观点，认为提取和对应类比体涉及相似性、结构和目的的约束。我们用以实现这些约束的并行满足的计算模型就用到了与上面讨论决策制定相似的机制。

解释 保罗·邱奇兰（Churchland 1989）提出解释应当理解为在分布式网络中编码的原型。理解为什么某一种鸟有一支长脖子可以通过激活表征天鹅的一群结点来实现，其中包含了对天鹅有长脖子这一原型化的预期。按照这一观点，推导最佳解释就成

为激活最恰当的原型。

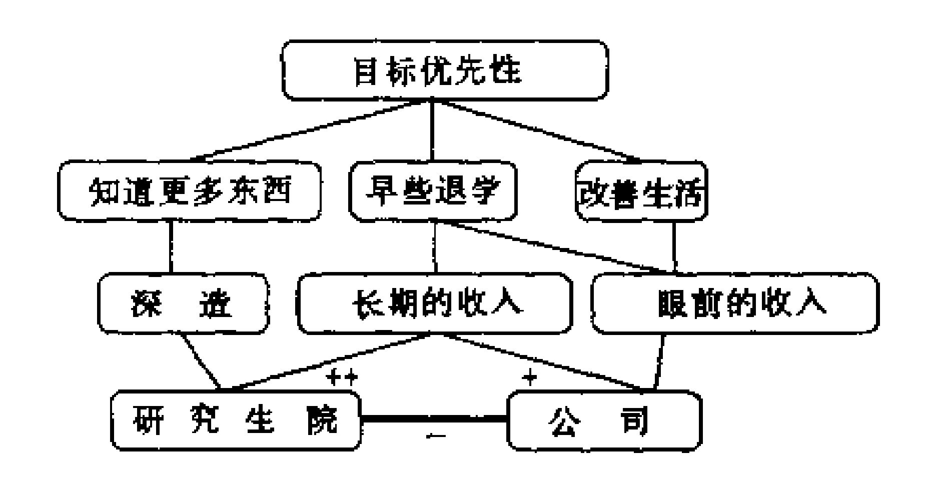


图 7.5 一个决策制定的约束网络

框盒代表单元，细线表示基于促进作用的正约束（对称的兴奋型联接），
带有减号的细线表示负约束（抑制型联接）。
“目标优先性”单元向相互竞争以获取激励的其他结点驱动激励。

如果采用局部式网络，推导最佳解释可以通过对解释的融贯性理论（Thagard 1989，1992）的模型化来实现。假设你正在咖啡馆里等待你的朋友弗雷德，但他却没有来。根据你对弗雷德的了解及你对其他同学的一般性知识，你可以对弗雷德为什么没有来提出各种假设。但你需要判断哪一个假设最具合理性。弗雷德可能是在学习，或者他约了别人去跳舞。弗雷德出现在图书馆这则额外的信息将会支持某一个特别的假设。图 7.6 所示是一个反映了用于 ECHO 程序的部分相关信息的网络，ECHO 是有关解释融贯理论的一个模型。单元用来代表各种证据材料，它们与一个能激活它们的特殊证据单元联接，然后将激励传输给其他单元。在代表弗雷德在图书馆和弗雷德去跳舞了这两个竞争性假说的单元之间有一条抑制性联接。选择最佳解释不仅涉及到某些特定假说

的证据，还与为什么该假设可能为真的解释有关。例如，弗雷德要去学习的动机是他想获得高分，而他去舞会的原因是他喜欢跳舞。网络最终的置定将对他的行为提供一个融贯一致的解释。在图7.6的网络中，网络最终置定为表示“弗雷德在学习”的单元被激活，因为它比它的竞争者——代表“弗雷德去舞会”的单元——有更多的激励源。

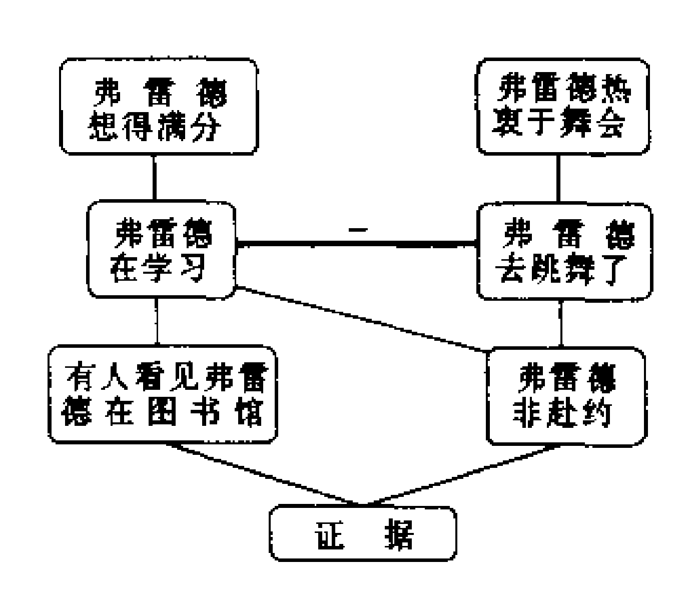


图7.6　对弗雷德为什么未赴约选择最佳解释的网络

细线是对称的兴奋型联接，而带减号的粗线是对称的抑制型联接

学　　习

在给定了联接主义网络的简单的结构后，可以有两种基本的方式进行学习：增加新的单元，或者改变单元之间联接的权重。目前的研究主要集中在第二种学习方法上。赫伯（Hebb 1949）提出了一个在生物学上可能的权重学习方法，他推测两个大脑细胞或系统同时被激活时，它们应该彼此联系起来。在真实的神经元中观察到了这种学习，而且可以以各种方式在计算上进行模拟。如

果单元（神经元）A 和单元 B 同时被激活，那么联接它们的权重就应当增加。例如，在一个局部式网络中有两个单元分别表示跳舞和舞会。如果这两个单元经常被同时激活那么它们之间的联接就会变得越来越强，体现出跳舞和舞会之间的联系。由于这种方式毋须任何教师告诉网络它的答案是对还是错，所以这种学习是没有监督的。

在前向式分布表征的网络中最为常见的学习使用了一种称为**反传**（backpropagation）的技术。图 7.7 显示了一个带有输入、隐藏和输出单元的简单的网络，用来学习校园中的社会性角色的原型。通过训练，网络应该能够对各种学生进行分类：给定一组特征激活输入层，在输出层应当能激活恰当的原型。例如，一个热衷于体育运动和舞会（输入层）的学生可以确定为"健将"（输出层）。通过下列步骤，使用反传算法来调整不同单元之间联接的权重来达到训练网络的目的（参见 Towell 和 Shavlik 1994；更详细的讨论见 Rumelhart 和 McClelland 1986）：

1. 在单元之间的联接上随机地分配权值。

2. 根据网络所需学习的特征激活输入单元。

3. 通过网络隐藏层再向输出层前向传播激励。

4. 通过计算输出单元的激励值与所要求的激励值之间的差异来确定错误。例如，如果**沉静**和**学习**刻苦的激励值激活了**"健将"**；这就是一个错误。

5. 将错误沿联接反传下来，以减少错误的方式改变权重。

6. 逐渐地，在向网络呈现了很多个示例之后，它便能正确地对不同类型的学生进行分类了。

反传模型在心理学和工程上都获得了许多成功应用，比起简单的确定性规则例如**如果**某人热衷于体育运动，**那么**他（她）是个"健将"适用面要大得多。经过反传训练的网络可以确定在输

入特征和输出特征之间较之规则更微妙的统计性联想关系。然而，作为一种入类学习的模型而言，反传仍有诸多不足之处。首先，它需要一个监督者来指出是否发生了错误，而很多的学习，比如语言，其学习过程中并无特别明显的监督纠正。其次，反传训练的速度很慢，训练一个简单的网络，需要提供成千上万示例。对某些类型的人类学习而言，大量的尝试是需要的，但有时人们仅从极少的例子中就能学到东西。

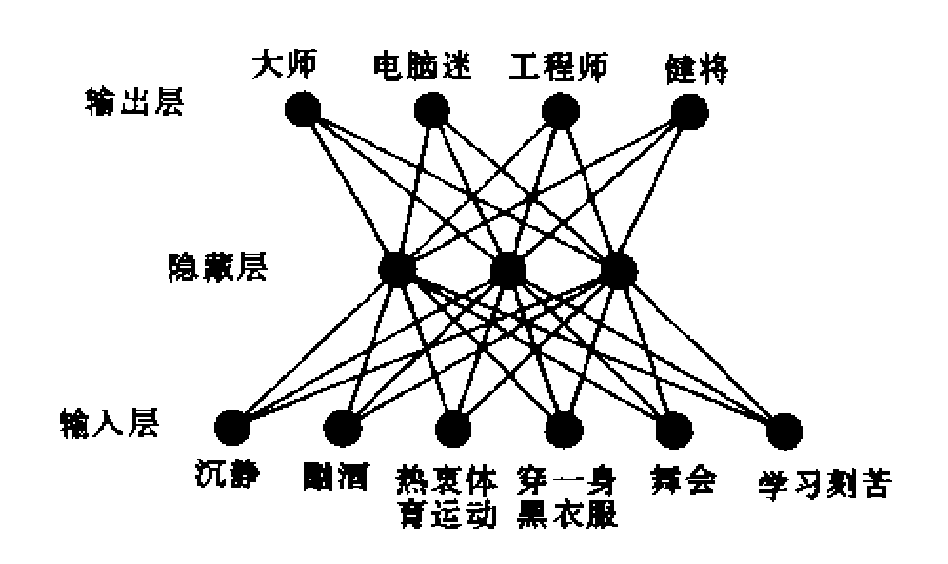


图 7.7　一个经训练后能对学生进行分类的网络

语　　言

早期的联接主义的语言模型，涉及语言的视觉知觉和听觉知觉。麦克兰和拉姆哈特（McClelland 和 Rumelhart 1981）揭示了语词认知可以如何理解为一个并行约束满足问题。假设把咖啡泼到本页上，致使有些字母被盖住了。但借助于尚可辨认的字母和整个的上下文，你可能仍然能推出被盖住的词语是什么。例如，在图 7.8 中，通过利用有关字母形状的已有信息和字母所在词语的整个上下文，有可能确定每一个单词中的模棱两可的字母。相互

联接的单元用来表征关于对什么字母呈现以及什么词语呈现的假设，而对网络的释放则可选出最佳的全局性解释。麦克兰和艾尔曼（McClelland 和 Elman 1986）研制了一个类似的用于话语知觉的模型。



图 7.8　借助上下文可以确定有不同字母选择的结构

联接主义网络除了可以确定模棱两可的字母和发声，还可以用来确定词语的意义。金茨（Kintsch 1988）提出了一个用于文本理解的“建构-整合”模型，可以解释诸如“bank”一词如何有时用来指一种金融机构，而有时指河岸。不同于第四章介绍过的概念观，意义不是封装进一个概念内，而是通过相互联接的元素建立在特定的上下文中。图 7.9 示意的是一个用于确定“bank”在特定上下文中适当含义的网络的一部分。激活哪一种解释依赖于输入信息如何影响各种单元和联接。

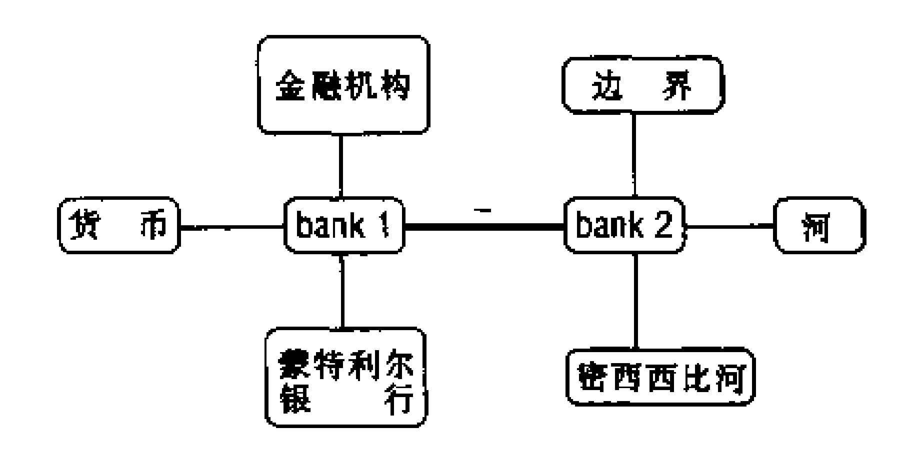


图 7.9　“bank”的意义由网络的激励流决定
细线是对称的兴奋型联接，粗线段是对称的抑制型联接

拉姆哈特和麦克兰（Rumelhart 和 McClelland 1986）研制了一个关于儿童怎样学习构造英语动词过去时而不需要形成清晰的规则的并行分布式处理的模型。对于为什么儿童错误地使用过去时如“goed”和“hitted”的一个解释是儿童已经形成了动词后面简单地添加“ed”的一条规则。错误出现在将这条规则过于普遍地应用到非规则动词上。但拉姆哈特和麦克兰揭示了一个经过训练的联接主义网络通过使用分布式表征而不是规则如何再现儿童所犯的错误。与此针锋相对，平克和普林斯（Pinker 和 Prince 1988）则指出联接主义网络在心理学上是不可能的，因为其生成过去时的方式与儿童所用的方式大相径庭。麦克温雷和莱因巴赫（MacWhinney 和 Leinbach 1991）回击说用一个重新设计的联接主义网络可以克服这些异议，而凌和马林诺夫（Ling 和 Marinov 1993）则用一非联接主义模型进行反击，认为联接主义模型不具备心理学的实在性。

心理学上的合理性

联接主义模型为很多的心理学现象提供了解释。上面介绍过的麦克兰和拉姆哈特（McClelland 和 Rumelhart 1981）的语词知觉模型就解释了好几个心理学实验的结果，例如，拉姆哈特和麦克兰（Rumelhart 和 McClelland 1982）介绍了一些心理学实验，能够证实他们的模型对于字母上下文呈现时间对一个词语的可知性所作的预测。麦克兰和艾尔曼（McClelland 和 Elman 1986）则描述了用他们的模型可以解释的各种言语知觉现象，比如在时间上的效应。同样，金茨（Kintsch 1988）的文本理解模型也通过学生确认各种类型语句的实验得到证实（Kintsch 等 1990）。

类比对应和提取的局部式联接主义模型不仅用于模拟以前的心理学实验的结果，还对新的实验提供了启发（Holyoak 和 Thagard 1995；Spellman 和 Holyoak 1993；Wharton 等 1994）。例如，

斯贝尔曼和赫尔约克（Spellman 和 Holyoak 1993）能够以计算机进行模拟的方式表明类比的目的对类比对应有影响。同样，为了测试我提出的解释性假设如何进行评价的联接主义模型，瑞德和马尔库斯-纽豪尔（Read 和 Marcus-Newhall 1993）及尚克和兰列（Schank 和 Ranney 1991，1992）设计了实验来比较人类被试与 ECHO 程序所作出的判断。昆达和萨伽德（Kunda 和 Thagard，1996）使用了一个简单的局部式联接主义模型对有关人们怎样形成对他人的印象的十来个实验进行了说明。

反传技术已用于模拟许多的心理学过程。例如，塞登柏格和麦克兰（Seidenberg 和 McClelland 1989）使用反传建立了一个视觉词语认知的模型，用于模拟人类行为的多个方面，包括词语如何在加工难度上有差异，新词汇如何发声，以及人们怎样从起步过渡到熟练的阅读。圣·约翰（St. John 1992）使用反传产生了一个模拟了文本理解的诸多方面的分布式表征模型。

神经学上的合理性

局部式联接主义网络在神经学上如何可能呢？本章中介绍的人工网络与大脑的结构有相似之处，它们都有简单的元素，相互之间进行激励或抑制。但真实的神经网络要复杂得多，有数以十亿计的神经元和数以万亿计的联接。不仅如此，真实的神经元也比人工网络中的单元要复杂得多，后者还只是在相互之间传递激励，而神经元有数十种神经递质，相互之间具有化学联接，所以不仅要以电学角度还须从化学角度来考虑大脑。真实的神经元在突触和非突触性质上都发生变化，这也超出了人工神经网络模型的范围。

在局部式表征中，每个单元都具有一个可指定的概念上的或命题上的解释，但大脑中的每个神经元却不具有这样的局部式解释。每个人工单元充其量只能看作代表了一个**神经元组群**，即一

群神经元共同工作来承担了一个进行加工处理的角色。将单元视为神经元组群而不是神经元还有助于解决单元和神经元之间的另一个差异：许多局部式网络在单元之间使用的是对称式联接，而联接神经元的突触却是单向的。但神经元通常具有能让它们相互影响的神经通路。此外与人工神经网络中的单元不同的是，真实的神经元与其他神经元之间的联接要么是兴奋型的，要么是抑制型的，而不可能是混合型的。毫无疑问，大脑肯定是将其表征分布到远比目前的局部式或分布式人工神经网络中数目多得多的神经元上面。

激励状态相似的神经元之间的突触联接会得到加强这一赫伯式学习方式，已通过对大脑的观察得到了证实，大脑还能通过对突触的调节实现其他类型的学习(Churchland 和 Sejnowski 1992，第五章)。然而，反传式学习还无法与科学家们在大脑中观察到的任何过程对应起来，实际的神经网络具有反传式网络的前向传播激励的特征，但对通过前向传递激励的通路用于反向传播错误校正的神经学机制仍一无所知。

所以说大多数的联接主义模型还是非常粗略地近似于真实神经元的行为。不过，在大脑与计算式心智之间这一类比到目前为止是非常富有成果的，而更为真实地类同于大脑的计算机模型也正在积极地研制中。

实践上的可应用性

联接主义模型的学习和效能在教育方面具有诱人的前景。亚当斯（Adams 1990）对于阅读所须具备的各种类型的知识进行了一番联接主义风格的描述。图 7.10 展示了在正字法、词语意义和一个词语所处的更广泛的上下文之间的相互关系。为了阅读一段文本，你需要处理词语当中的字母并且同时考虑到词语的意义和上下文。按本章的术语来说，阅读是一种并行约束满足，须同时

涉及的约束包括拼写、意义和上下文。在阅读教学中，任何忽略这些约束的狭隘方式，比如忽视发音或是忽视意义和上下文，都会增加阅读学习的困难。

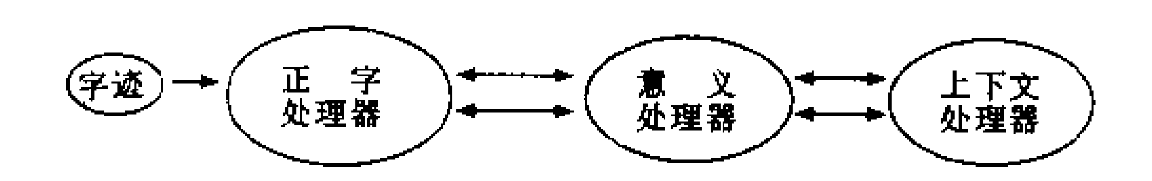


图 7.10 阅读中所需的多个处理器

（Adams 1990，第 138 页），也可见 Seidenberg 和 McClelland 1989

设计自然离不开并行约束满足。例如，一位建筑师对一栋大楼的设计要考虑到很多方面的约束，如成本、大楼的用途、楼房所处的环境以及美学上的考虑。反传技术已经用于辅助工程师们预测大楼所需材料的压力和张力（Allen 1992）。

联接主义模型在智能系统中得到了广泛应用。反传算法有许多的工程应用，例如用在训练网络认别炸弹、水下目标和笔迹。一家银行训练了一个人工神经网络来确定它的哪些客户可能会拖欠贷款，另有一些网络用于解释医学测试的结果并预测疾病的发生。维得鲁、拉姆哈特和雷尔（Widrow，Rumelhart 和 Lehr 1994）对神经网络在工业上的最新应用进行了综述。

小　　结

由简单结点和联接构成的联接主义网络对于理解涉及并行约束满足的心理学过程来说是非常有用的，这些心理过程包括了视觉、决策制定、解释的选择和语言理解中的意义构造等诸多方面。联接主义模型可以借助赫伯学习和反传等方法来模拟学习。联接主义的解释程式是：

解释目标

为什么人们会具有某种特定的智能行为？

解释模式

人们具有由兴奋型和抑制型联接相互联在一起的简单处理单元构成的表征。

人们具有通过联接在单元之间传递激励及修正联接的处理过程。

将传播激励和学习的过程应用于单元上便产生了行为。

对各种心理学实验的模拟展示了联接主义模型在心理学上的相关性，不过它们与实际的神经网络只有粗略的近似性。

讨论题

1. 局部式表征和分布式表征有何区别？
2. 人工神经网络的单元与实际的神经元有何区别？
3. 联接主义对心理学现象的解释与基于规则的解释有何不同？
4. 哪些心理学现象可以最自然地用联接主义的方式来解释？
5. 哪些心理学现象最难于用联接主义的方式来解释？

进一步的推荐读物

神经网络模型的入门读物有 Bechtel 和 Abrahamsen 1991，Churchland 和 Sejnowski 1992，Levine 1991 及 Rumelhart 和 McClelland 1986。McClelland 和 Rumelhart 1989 提供了如何建构你自己的模型的详尽指导。Anderson 和 Rosenfeld 1988 收录了部分有关神经网络的经典文献。Hinton 1990 提供了对联接主义学习程序的综述。认知神经科学的入门书，有 Kosslyn 和

Koenig 1992。

各种较新的联接主义模型，参见 Bechtel 1993，Pollack 1990 和 Smolensky 1990。Churchland 1995 提供了一门可读性强的关于回复式网络的导论书籍，回复式网络不仅包含前向式联接，还含有反馈式联接。Prince 和 Smolensky（即将出版）提出了一种部分基于联接主义思想的关于生成式语法的新路线。

备　　注

Boden 1988 通过与儿童相互传递消息进行类比来对各种类型的网络进行解释。

第四章谈到的概念之间的那种传播激励比起本章中所讨论的要狭隘得多，后者不仅包括了兴奋型联接机制，而且还包含抑制型的，并且还含有不表达任何完整概念的隐单元。

为了计算一个联接主义网络中单元的激励值，每个单元选出一个初始激励值，在开始学习后则重复循环。有很多种方式来完成此事。在一个常见的技术中，在每一个循环时单元 j 的激励值 a_j，按照下面的方程式来更新：

$$a_j(t+1) = a_j(t)(1-d) + \begin{cases} net_j(\max - a_j(t)) & \text{如果 } net_j > 0 \\ net_j(a_j(t) - \min) & \text{其他情况。} \end{cases}$$

这里 d 是一个衰减系数（每个单元的系数在每次循环后要减小），min 是最小激励值（-1），max 是最大激励值（1）。根据单元 i 与 j 之间的联接权重 w_{ij}，我们可以用下式计算出净权重 net_j：

$$net_j = \sum_i W_{ij} a_i(t)。$$

第八章　回顾与评价

认知科学与摇滚音乐可能说是同龄人，它们都由各个不同的源流会聚而成，在50年代中期问世。与摇滚音乐一样，随着新思想和新技术的发展，认知科学在很多方面发生了变化。本章简要地总结认知科学所取得的成就，对第二章至第七章介绍过的六种基本路线的表征力和计算力进行比较和评价。此外还简要介绍对CRUM即对心智的计算-表征理解所提出的一系列重要挑战。

认知科学的成就

与40年前行为主义占据统治地位时相比，现在对问题求解、学习和语言的科学理解无疑要深入和复杂得多了。我们知道怎样设计能进行逻辑推导的复杂系统，基于规则和基于概念的系统也已成功地对问题求解和语言使用的各种方面进行了模型化。在过去的15年间，借助于心理学实验与计算机建模的相互结合，对类比性思维的理解也得到了增长，表象也从地处科学探索的边缘转移为心理学、神经学和计算研究的重要论题。联接主义的学习和并行约束满足模型为很多的心理学现象提供了解释。

认知科学未竟的目标之一是为全方位的心理现象提供一个统一的理论，就像进化论和遗传学统一生物学现象，相对论和量子力学统一物理学理论那样。不同的认知科学家提出心智是一个逻辑系统、基于规则的系统、基于概念的系统、基于类比的系统、基于表象的系统以及联接主义式的系统。本书的看法是对于“心智是什么样的系统?”这一期末考试试题，目前的最佳答案是：“上面得到的都是”。心智是一个超乎寻常的复杂系统，有着各种各样

不同种类的思维活动。

在第二章至第七章中介绍的CRUM的不同路线都致力于抓住心智的不同方面。表8.1给出了对这些不同方式的小结及它们的理论应用。在认知科学探索的初级阶段，理论的多样性与其说是一个缺陷，不如说是一个令人称心的特点。当然，我们也期待着认知科学的牛顿、达尔文或爱因斯坦的面世，为认知科学提供一个简单的、统一的理论，整合目前所有的见解。心智的复杂性和多样性可能使得这样的理论有些可望而不可及，但即使还没有这样的大一统理论，对心智的理解仍然可以取得进步。认知科学的前提之一便是认知科学的进步要求的不是某一特定学科的研究者的孤军奋战，集成式的、跨学科的努力对于理解心智的本质是至关重要的。

表 8.1 对各种计算方式的理论应用的回顾

	表 征	问题求解	学 习	语 言
逻 辑	命 题 算 子 谓 词 量 词	演 绎 概 率	概 括 逆 推	解 析
规 则	如果-那么	搜 索 前向链 反向链	组 块 概 括 逆 推	语 法 发 音 拼 写
概 念	带有槽的框架 程 式 脚 本	匹 配 继 承 激励传播	从案例中抽取 概念组合	语 典 语义学
类 比	目标和源 因果关系	提 取 匹 配 修改采纳	存贮记忆 形成程式	隐 喻
表 象	视觉、运动等	匹 配 操 作	表象式训练	表象程式
联 接	单元和链路	并行约束满足	反 传 权重的调节	消除二义性 发 音

认知科学还在教育、设计和智能系统等方面产生了实质性的应用，对心智不同方式的计算-表征理解对应用性思维的不同方面

都带来了启示。例如，我们已经看到，基于规则的系统和类比模型有助于理解学生如何求解问题，联接主义的并行约束满足对阅读教学有重要的应用。设计要求不同的认知过程的参与，从演绎推理到表象。模仿人类能力的智能系统利用了各种类型的表征和加工过程，尤其是基于规则的、类比的（基于案例）和联接主义式的反传系统。

表 8.2 认知科学的实际应用

	教　育	设　计	系　统
逻　辑	批判性思考	编　　码	逻辑编程
规　则	算　　法 技能的获取	人机交互	大多数的专家系统
类　比	问题求解	基于案例的设计	基于案例的专家系统
表　象	视觉问题求解	图　　表	部分专家系统
联　接	阅　　读	约束满足	专家系统训练

比较性评价

为了对六种表征与计算的不同方式有一个更深的了解，我们可以继续采用表征力、计算力、心理学上的合理性、神经学上的合理性和实践上的可应用性的标准对它们各自的优缺点进行比较评价，这一比较使我们更确信当前没有任何一种路线能成为整个认知科学的基础。

表征力

我们已见到形式逻辑具有相当可观的表征力，借助算子如“非”和“或”以及量词“所有的”和“有些”能够生成复杂的命题。局限于规则、概念、类比、表象或联接的计算机模型在表达

诸如“没有一个学生导师应当对他们的某些学生的困难和焦虑负责任”这样的复杂命题时都颇为棘手。即便如此，形式逻辑也无法抓住自然语言的所有精妙之处，所以我们不得不承认目前的计算模型所具有的表征力尚无法刻画人类思维的全貌。

联接主义模型胜过语言式表征之处在于具有更大的灵活性来反映更为广泛的感知经验。单元的激励模式可以表征许多味觉和嗅觉经验，而语言式表征对此只能求得近似。而另一方面，简单的类神经元单元如何能表达那些能自然地包含在基于逻辑、规则或类比的计算模型中的复杂关系，这是联接主义无法回避的挑战。由数十亿个神经元组成的大脑终究是能产生语言的，我们也知道一个基于相互联接的单元的系统能产生出复杂的推理，但实现这一步飞跃要求联接主义模型比目前拥有更强的表征力。

基于规则的计算系统放弃了逻辑所提供的表征上的丰富性，而只采用了在计算上更有优势的简化的**如果-那么**规则。与形式逻辑中的命题一样，规则既简洁而又在表达上相互独立，相反，概念、类比和表象则把信息束集为有组织的结构。一个概念是关于某一类事物的一个信息封装，而一个类比则将关于一个情境的信息聚拢在一起，表象提供的是一种特殊的封装方式，与视觉这样的感知功能紧密地结合在一起。一个视觉表象生动地将相互联系的信息拴在一起，而这用语言来表达却很困难。

总之，一个统一的心理表征理论必须在它提出的结构中具有：(1) 表象和联接所具有的感知丰富性；(2) 概念、类比和表象所具有的组织能力；(3) 规则和形式逻辑的命题所具备的语言表现力。

计 算 力

在研制计算模型时，我们需要考虑到速度、灵活性及抽象的计算能力。实现计算有很多种方式，但对于认知科学我们需要的

是能够满足心理学上的合理性和实际可应用性的足够的速度和灵活性。逻辑演绎推理可以做得简洁优雅，但侧重于启发式搜索的基于规则的系统在许多领域却有更胜一筹的表现。规则系统的有效性使得纽威尔（Newell 1990）这样的理论家会提出一种基于规则的统一的认知理论。其他的计算机制还包括应用了概念、类比和表象的整体性结构的匹配。基于概念的系统和联接主义模型能实现不同的传播激励。虽然很多的人类问题求解可以在一个基于规则的系统中构造成启发式搜索，但仍有很多的问题解决更适合于像程式运用、类比对应和并行约束满足等过程来实现。

同样，人类的学习也不限于一种单一的机制如基于规则的组块。一个全面的理论要考虑到从事例以及其他的规则和概念的组合中学习规则和概念。它应当既包括快捷的、一次性的学习，如人们逆推式形成新的假设，也应包括缓慢的、多次性的学习，如儿童学会保持身体平衡。基于规则的组块过程和联接主义的权值调节都是强有力的学习机制，但都不能够概括人类学习能力的全部。

虽然不同的研究路线对语言的各个方面都带来了可观的新见解，但对于语言的学习和使用，认知科学仍缺乏一个综合性的理论。例如，对于语法和发声的某些问题，可以由规则加以描述，但基于规则的方法对于理解词典的本质或者隐喻在语言生成和理解中的角色都无甚助益。因而，除规则以外，语言似乎还与概念、类比和表象有关，也许联接主义能逐步在一个单一系统里对语言的所有这些方面提供一种侧重神经学的说明，不过，尽管目前联接主义的学习和并行约束满足模型对于诸如词语意义的确定等语言学应用相当成功，但一个综合性的联接主义语言理论还尚未出现。

心理学上的合理性

上述六种表征和计算的基本路线对于心理学实验和计算模型

的研制都提供了启示。这些实验所针对的诸多富有争议的论题还将继续争论下去。三段论和其他类型的逻辑推理是通过运用逻辑规则，还是一些更具体的方法如心理模型来完成的？人们学会生成英语动词过去时这一过程是用联接主义模型还是用规则来描述更好？实验心理学 90 年来的发展使得许多的心理学现象都要求一个一般化的心智理论来给予说明，但目前的情况是不同的实验结果与不同的表征理论相互吻合。基于规则的模型能较好地应用到诸如连子棋这样的认知任务，但对于类比式的问题求解的案例却不能告诉我们更多东西。实验对表象在人类思维中的重要性给予了支持，但仍有许多现象看来与表象无关。虽然在第七章中介绍的联接主义模型对于说明多种心理学现象是成功的，但如果说其他类型的模型都不必要了，条件还不成熟。联接主义模型比较能适用的认知任务是那些能够理解为累积式的学习和并行约束满足的问题，但单元和约束的生成可能需要基于规则的或其他类型的机制，而这是目前的联结主义模型还未触及到的。

若有一个统一的认知理论能够说明目前已观察到的所有心理学现象，这无疑是一件妙不可言的事情。但进步也可以是局部式的，研究针对特定心理学现象的特定的表征和计算理论，在研制丰富、能说明心理学实验中所观察到的人类表现的计算模型的过程中，认知科学取得了实质性的进步。不同类型的表征和思维如何能够相互结合起来，对这个问题的回答无疑还需要更多的实验、更多的整合式的模型（第十二章中将要讨论）。

神经学上的合理性

我们已经看到对心理表象和大脑视觉系统之间的联系而言，已有相当可观的神经学上的证据。联接主义模型从人工神经网络与大脑之间的类比中也得到了一定的神经学合理性，虽然目前的联接主义思想与实际大脑如何工作相比还只是非常粗略的近似。

规则、概念和类比在神经学证据上的匮乏并不意味着这些表征在神经学上是不可能的，因为目前的方法和技术的水平还不足以对它们的角色作出评判。可以观察大脑活动状况的新的扫描技术的出现，使得认知神经科学成为认知科学发展最快的一部分。如果本书在2000年时出修订版，变化最大的部分很可能就是对神经学合理性的讨论了。

实际上的可应用性

构造一个统一的认知理论要求将不同的认知科学家关于心智本质上是一个基于规则的系统或是一个联接主义的系统等相互冲突的主张调和在一起。但是要实现一些实践目标，如改进教育、设计和智能系统，则可以要用较为零敲碎打的方式，有选择地从认知科学研究的不同路线中选用合适的内容。

从根本上说，认知科学对于教育来说就如同生物学对于医学：是实际治疗技术的理论基础。规则、概念（程式）和类比构成的心智概念业已对理解人们怎样解题做出了贡献。图表在很多领域中的应用，无疑也是表象与问题求解相联系的证据；联接主义思想的观念对教育理论和实践的影响才刚刚开始，从并行约束满足的角度来理解阅读过程对于教学的改进有启发作用。

目前，重视规则、概念、类比和表象在创造性设计的角色极大地推进了对设计过程的理解。具有工业应用的大多数专家系统都是基于规则的系统，而基于案例（类比）的系统和联接主义系统也正日益显现出应用的前景。一个希望开发出智能系统的经理应当细致地了解所要完成任务的本质以及可供使用的知识，审慎地考虑哪些种类的表征和计算最为恰当。

一些认知科学某种特定路线的拥护者总是大胆地宣称心智是一个基于规则的系统，或者是一个联接主义式的系统，如此等等。事实上目前所有的表征和计算的理论都是优点和缺陷并存，这说

明各种路线的相互结合是必要的（参见第十二章）。而CRUM的一些批评者也指出，所有这些计算方法对于告诉我们心智是什么而言都有其内在的局限性。

对认知科学的挑战

通过对心智的计算-表征理解的主要研究路线的回顾，我们可以看到CRUM能够对人类问题求解、学习和语言的本质作出很多解释。虽然CRUM对于说明心智的本质取得令人可喜的成功，持怀疑态度的人们认为它带来了根本性的误导，忽视了思维的关键性的方面，比如意识和情绪经验，第九章至第十一章将讨论六种对CRUM的重要的挑战。

1. **情绪的挑战**：CRUM忽视了情绪在人类思维中的重要作用。
2. **意识的挑战**：CRUM忽视了意识在人类思维中的重要性。
3. **物质环境挑战**：CRUM忽视了物质环境在人类思维中的重要作用。
4. **社会性挑战**：人类思维有其固有的社会性，而CRUM没有考虑到这一点。
5. **动力学系统挑战**：心智是一个动力学系统，而不是计算系统。
6. **数学上的挑战**：数学上的结果表明人类思维在标准的意义上不是计算的，所以大脑是以一种完全不同的方式运作的，可能是一种量子计算机。

这些挑战对CRUM提出了严重的质疑并已危及整个认知科学的事业。对此有四种可能的回应：

1. **否认**构成这些挑战的基础的那些主张。

2. **扩展** CRUM 使其能解决这些挑战所提出的问题，增加新的计算和表征观念。

3. 用非计算、非表征的观念**补充** CRUM 使得 CRUM 能够面对这些挑战。

4. **放弃** CRUM。

我将会论证这些挑战提出的理由都不足以使我们放弃 CRUM，不过，这些挑战也表明 CRUM 需要加以扩展和补充，特别要能够与生物学和社会性因素进行整合。补充不同于扩展在于它要引入超越计算-表征解释模式的概念和假设。

小　　结

对心智的计算-表征理解对于理论理解和实际应用都作出了很大贡献。但对于人类的认知能力，还没有一种单一的方式能成为最具威力的解释，不同的方式在表征和计算上各有其优劣。各种不同方式在心理学上都有一定的合理性，能够成功地为不同类型的思维提供模型。但 CRUM 面临着挑战，它忽视了心智的许多重要的方面。

讨论题

1. 认知科学最突出的成就是什么？在哪些方向上它仍有很长的路要走？

2. 是什么对一个统一的心智理论构成了障碍？我们会具有一个统一的心智理论吗？

我们需要吗？

第九章　情绪与意识

心-身问题

本章讨论情绪与意识对CRUM构成的挑战，由于这些挑战引发了一些有关心与身的本质的重要的一般性问题，因而首先简述关于心与身关系的哲学见解会对我们有所助益。

关于人（person）是什么的常识性观点认为人由两个部分所构成：身体（body）和心灵（mind），这种观点被称为**二元论**，它假定每个人都由两种不同的基本材料所构成，一种是物质材料而另一种是心理的或精神的材料。任何一种认为人死后灵魂依然存在的宗教观点都是二元论的；因为灵魂只有是某种非物质的东西才能在肉体死亡后继续存活。虽说二元论可能是关于心灵存在的最为普遍接受的观点，它在哲学上却很成问题。我们有什么证据说心灵独立于肉体呢？如果说心灵与身体是两种不同材料组成的，它们又如何相互作用呢？二元论把心灵弄成了逃离于科学研究之外的、本质上神秘的实体。

与二元论针锋相对，**唯物论**认为心灵不是由异于构成身体的物质材料的另一种材料构成的。哲学家们提出了数种不同类型的唯物论观点，**还原式唯物主义**认为每一种心理状态，比如意识到油炸面饼圈的香味，是一种大脑的物理状态，因此，心理状态都可以还原为物理状态。更激进的**排除式唯物主义**则认为我们不应将我们所有的心理经验与大脑事件等同起来，因为我们关于心灵的常识性观念可能在根本上是错误的，进而言之，随着神经科学的发展，我们可以指望发展出关于心智的更为丰富的理论，取代

并排除掉诸如意识和信念等常识性观念。

不论是还原式的还是排除式的唯物主义都假定对心智的理解在根本上依赖于对大脑的理解。不过，对心智的计算式研究则通常与另一种被称为**功能主义**的观点相联姻。功能主义认为心理状态并不必然是大脑状态，而是通过因果联系相互关联的物理状态，而这样的因果联系可以由各种不同类型的物质来承载，例如，一台智能机器人可以看作具有心理状态，即便它的思维依赖于硅芯片而不是生物神经元。同样，我们也有可能遭遇上来自别的星球的有智能的异类，其心理能力依赖于与人脑大不相同的生物结构。

这四种观点——二元论、还原式唯物论、排除式唯物论和功能主义——是目前的心智哲学中曾经走红的哲学主张。不过，我要说的是，对情绪和意识的考虑支持另一种立场——综合式唯物主义——主张计算、神经生物学和意识经验的理论综合。

情绪的挑战

在电视连续剧《星际旅行》中表达了人们对于情绪的一种常见观点：人类思维内在地受到情绪支配，因而是非理性。来自祝融星（Vulcan）的外星人斯波克先生不受情绪影响，因而其思维表现出超人的逻辑性。同样，在《下一代》一集中，机器人德特（Data）具有一个计算机大脑，可以不受情绪影响进行运作。直到续集《世世代代》中，他装上了一块情绪芯片，从而致使其行为变得反复无常。

而在神经科学家安托尼奥·达马西奥(Damasio 1994)最近的论著中，为我们展现了有关情绪在思维中的角色的另一种景象。达马西奥介绍了一批大脑有损伤的病人，他们的新皮层与扁桃核之间的连接被切断。新皮层是高层思维活动的大脑区域而扁桃核则是靠近脑干的部分，情绪活动主要发生在这里。这些大脑有损伤的病人就变成了人类的斯波克和德特，通常认为，高层的皮层功

能显然应当不受情绪的影响。然而，让人吃惊的是，这些病人丝毫也不像超常的逻辑家。达马西奥介绍说他们在语言和数学能力上没有任何缺陷，但他们的日常生活能力却受到了很大的限制。例如，有位病人可以长时间地来回比较不同餐馆的优缺点，却绝不进去用餐。另外的一些病人在他们的社会交往中变得极不负责任，因而无法保持与他人的关系或工作职位。在达马西奥看来，情绪对人类的思维和行为来说是一个至关重要且极有影响的方面。

不管我们将情绪视为人类思维的一个积极的还是消极的部分，情绪对心智的计算观确实构成了严重的挑战。在第二章至第七章中介绍的表征与计算路线的认知研究对于情绪都没什么说道，而且这种计算路线似乎与对情绪的重视是相对立的。那些相信思维类同于在没有情绪芯片的计算机上执行的计算操作的人会倾向于否认情绪与思维有多大关系。

对情绪挑战的回应

否　认

一个 CRUM 支持者可以坚持心智-计算机类比的中心地位，强调情绪处于人类思维的边缘，只是作为我们从更低等的生物进化而来而存留下来的残留物。但是这种简单地否认情绪对CRUM的挑战的策略，其代价是忽视了人类思维中较为普遍的一个方面。欧特莱和邓肯（Oatley 和 Duncan 1994）在实验中要求受试者记录下在规定的时间间隔内的思维活动及情绪，在大多数情况下受试者都有情绪上的体验。文学作品和我们的日常经验也为情绪在人类思维和行为中强有力的作用提供了丰富的例证。让我们来看看怎样扩展和补充CRUM 以迎接情绪的挑战。

扩展 CRUM

各种有关心智的计算理论都将人的思维视为采用不同的策略

以解决如何达到其目标的问题。情绪理论家如凯思·欧特莱(Oatley 1992)已阐明了人类的基本情绪如何与目标的完成密切相关。当其目标得以实现时人们会很高兴，反之则会感到沮丧。如果你在考试或者求职面试中发挥出色，从对你的职业或社会目标的满足中快乐油然而生，而这些目标的失败则会产生失望和沮丧。当某事对你实现目标构成挫折，比如某人占用了你的停车位，你会感到愤怒。当你生存目标受到威胁时你会体验到恐惧，比如一辆大卡车急刹车后仍向你冲过来。当你从你的食品中吃出了只苍蝇，恶心反映了对你进食目标的破坏。由此我们可以看到情绪涉及到对一个人总体问题求解情境的非常一般性的表征。

为什么要采用这种一般性的表征？为什么不用语言的或视觉的表征来反映当前的目标实现的状态？欧特莱指出人类的问题求解通常是非常的复杂，涉及要实现的多个相互冲突的目标，迅速变化中的环境以及丰富的社会交往。情绪可以对你的问题求解的情境提供了一个总体性的**评估**，对其后的思维过程产生两种重要的作用。你所处情境的某些特定方面对于实现你的目标是极端重要的，对此的评价有助于聚焦于这些方面，将你有限的认知资源集中使用到相关的部分上。不仅如此，情绪还能对应采取的行动提供预备，确保你警觉地处理你面临的问题求解情境，而不至于迷失在思索的迷雾中(Frijda，1986)。因此，情绪不只是人类思维的附带发生而令人厌烦的属性，而是具有与评价、聚焦和行动相关的重要的认知功能。

由于情绪在人类思维和行动中担任的角色，对人们行为原因的解释经常要求我们提到情绪状态："他愤怒之极，一拳砸在墙上"，"她的脸上整天挂着笑容，因为她被医学院录取而欣喜若狂"。有时候，基于情绪的解释会超出言语的表达，当我们理解他人的情绪时可以将自身置于他人所处的情绪中，并体验一种接近于他人感受的情绪。这种类型的理解被称为**移情**。它基于类比思维，我们在他人所处的情境与我们自身经历过的情境之间建立一

种对应，从而在我们身上产生出他人正在体验的某种情绪的意象(Barnes 和 Thagard，印刷中)。

情绪的聚焦作用可以影响到学习。在自然系统或计算系统中，有关学习的一个问题在于在哪些方面须集中精力以及在何处运用各种可行的学习机制。将适当的情绪表征与其他表征联系起来可以提供学习的聚焦点：如果《星际旅行》让你激动，你很可能会在它的驱动下去学习很多有关角色以及画面制作的知识。

情绪在人类思维中的重要性由此提示我们扩展前面各章中问题求解的计算观，加进诸如**快乐**和**悲伤**等新的表征结构，并与对目标实现的一般性状态的评价及对相应情绪表征的激活联系起来。情绪表征的激活就可以在计算上影响推理和决策的制定。例如，**恐惧**结构的激活就会驱动诸如逃逸的行动，其中使用的某条规则不是导致推理而是直接改变行为。

也许编写一个基于规则的系统就可以产生上述基于情绪的解释。同样，建造一个基于概念的程序，将情境与一个诸如**快乐**的情绪的原型结构进行匹配，这也是可能的。类比的计算模型可以用来模拟移情，其中从源类比体到目标类比体的转换不是语言式或图像式的表征，而是关于情绪经验的表征。一个联接主义式的系统可以用一个处理单元(或一组单元)来代表情绪结点的激活，比如**快乐**。在这样一个并行约束系统中由这样一个结点来完成的对目标实现的总体评估就显得尤为自然，毋须用明确的语言形式来表达目标是否得以实现。

因此我们可以模拟情绪对人类问题求解和学习产生作用的一些方面，但是模拟仍缺少了人类思维中某些核心的内容。激活**快乐**结构对于体验快乐而言不过是一个过于苍白无力的替代。二元论者会说要感受到快乐，你必须具有一个在本质上不同于计算机构成的由另一种材料组成的心。即便赞赏 CRUM 的唯物论者也能看出我刚才提出的对快乐的模拟较之对问题求解的模拟也要弱得多。当我们选择一个使用规则、类比或其他的表征来求解问题

的程序时，很容易认为计算机确实在解决问题。相反，对情绪的模拟就完全不同于让计算机具有了情绪，这就像在计算机上模拟台风完全不同于真正的呼风唤雨一样。

补充 CRUM

对人类来说，情绪经验不仅只与大脑紧密相关，还与整个身体的物理反应有密切关联。例如，愤怒就与血压的升高及胃酸分泌的增多有联系。不同的情绪体验也似乎与脸部血流的不同模式有关。因而人类的情绪经验与我们身体的不同部位有着密切的联系，而不仅仅与抽象思维过程有关。因此给《星际旅行》中的机器人德特添上一块计算机芯片而不赋予他人的身体，就认为他也可以具有人类的情绪经验，这是毫无道理的。

理解人类情绪看来至少要从三个方面着眼。首先，要从认知角度考察对我们的问题求解情境的评估，这与 CRUM 是联系在一起的，但仅有 CRUM 不足以完全刻画情绪是如何有助于评估、聚焦和行动的。其次，像快乐这样的心理经验更难以置于 CRUM 的解释范围，尽管具有经验过的表征而不是加工出来的表征在计算上有较多的优越性。举一个极端的例子，假如有一条响尾蛇在你面前爬，对这一事实的表征就不应当只是你心中激活的成千上万条表征中的一条：你需要把你的注意力直接集中到蛇上面，并采取行动避开它。不仅如此，你的情绪经验及其他的表征应该能使你更好地将情况告诉其他人，以利于群体规划（Oatley 和 Larocque 1995）。

理解情绪的第三个方面是大脑在认知与身体体验之间相互联系中的角色，这方面正在不断地产生出新的见解。达马西奥（Damasio 1994）报告的情绪缺损的病人是新皮层与扁桃核之间的连接被切断了，后者被勒都克斯（LeDoux 1993）称为一台“情绪计算机”。扁桃核有许多来自感知过程的输入及到运动系统的输出端，它的作用似乎是一个对身体的一般性状况进行评估并促发适

当行动的一个评估中心。对人类情绪的理解因而不能局限于抽象的计算研究，还要了解人的身体和大脑如何采用特定的机制来产生意识经验，并利用情绪来影响评估、聚焦和行动。CRUM 要由生物学方面的研究来加以补充，要探讨在人类身上，而不是在抽象的计算系统中，情绪在思维活动中所扮演的特殊角色。

对情绪的生物学基础的另一个视角来自于对神经递质以及对其有影响的药物的研究。真实的神经元远比第七章里介绍的联接主义模型中的人工神经元要复杂得多。它们相互之间不仅传递激励，在一个神经元对另一个神经元的作用过程中至少有 50 多种神经递质介入。5-羟色胺（血管收缩素）是影响多种行为的一种极为重要的神经递质。数以百万计的人在医生的建议下服用了帕罗扎克以治疗从抑制症、强迫症到易盛怒等心理问题（Kramer 1993）。尽管副作用很明显，但很多人报告说治疗的结果对于他们的情绪状况有实质性的改善。帕罗扎克能促进对 5-羟色胺的吸收，使得神经递质能容易地为一些特定的神经元接受。其他像多巴宁这样的神经递质与堕入爱河这样的情绪变化有关系。要深入了解情绪在人类思维中所担任的角色必须注意到对不同的神经递质的心理效应的研究中不断增长的新知识。

对心-身问题的传统观点对于情绪与思维的关系的回答都有太大的局限性。二元论强调有意识的情绪经验而忽视了不断增长着的对于与情绪变化相关的生理变化的了解。还原式的和排除式的唯物论过窄地局限于大脑和其他生理物质，却未能重视情绪的经验和计算等方面的情况。功能主义既忽视了意识经验也忽视了人类情绪的生理基础，即便机器人和外星人可以具有在很多方面与我们相似的智能思维，他们却不一定具有同样的情绪经验，因为这可能源自于我们特有的内分泌系统和神经系统。

我们需要发展一种既能认真看待情绪的经验内容和生理基础，而又兼顾其计算角色的**综合式**的唯物论。这种发展要求我们同时扩展和补充CRUM。扩展意味着要引入新的表征来刻画人类

多种情绪的某些基本的方面，同时要用新的计算程序来描述情绪表征在人类思维中的角色。更重要的是，CRUM 需要加以补充以容纳情绪的那些不可排除的经验和生理方面的特性。如果这些扩展和补充能够得以成功施行，那就没有必要放弃在解释非情绪式的问题求解和学习等方面 CRUM 已取得的相当的成功。在已达到 CRUM 需要更多地靠向生物学这一认识后，现在让我们从更一般性的层面上来考察一下经验与计算之间的关系。

意识的挑战

我们的意识经验不但包括快乐、悲哀和恐惧这样的情绪，还包括由我们的视、听、触、味和嗅的感知而产生的知觉。不仅如此，我们还能够意识到我们的信念和愿望，眼下我意识到我的信念是我正在写关于意识的这一章，而我的愿望是力争在午饭前完成这一节。我们还能意识到各种身体上的感知，比如疼痛和使用锤子钉钉子的感觉。哲学家用**感知特性**（qualia）（其单数是 quale）一词来指称意识经验。我们的感知特性具有一种整体性，构成了连续的意识流的各部分。

从 CRUM 的角度看，意识确实是一个谜。按 CRUM 的观点，思维是由在表征之上运行计算程序来实现的，而我们讨论过的六种主要的方式都对意识的特定角色闭口不语，那些针对规则激发、概念激活、类比和并行约束满足的程序似乎都是在意识水平之下运作的：我们自己不能直接意识到执行这些程序的算法。表象显然与意识的联系较为密切，因为我们能意识得到图形的形象，但对心理表象的计算和神经学说明却很少提及这种意识。

CRUM 不仅仅是倾向于忽视意识，如果意识从根本上超出了计算解释的范围，这对 CRUM 来说就是不得已的选择了。我们还没有任何理由认定目前我们用来模拟人类思维的计算机就具有意识，这样一来有关问题求解、学习和语言的计算模型也许应坚持

意识大体上与人类思维是不相关的。二元论者和其他对认知科学持批评态度的人由此得出结论，认为CRUM从根本上说不适合成为一个关于心智的理论。他们宣称任何一个忽视了意识这一核心心理现象的心智理论都应予以摒弃。

对意识挑战的回应

否　认

一些CRUM的支持者试图否认意识在人们心理生活中的核心地位，排除式唯物论者有时就视意识是一种前科学概念，随着科学的发展要被清除出去。按照这种观点，意识只不过是民间心理学（folk psychology）理论的一个部分，随着认知科学的不断发展最终被遗弃，正如有关女巫和妖魔的理论被现代人遗弃一样。很可能就根本不存在感知特性，就像不存在叫作**卡路里**的物质，19世纪以前的物理学家曾认为热是由卡路里流构成的，而当将热视为分子运动的热力学理论出现以后，卡路里这一概念便不复存在了。

然而，卡路里是不存在，女巫和妖魔也不存在，但要否定我们人人都具有的经验却并非易事。你知道比萨饼是何味道，汽油味是什么样的，一朵鲜花看上去如何，而被你所关心的人拒绝是怎样的感受。毫无疑问随着认知科学的发展，民间心理学的许多内容要发生变化。在第二章至第七章中介绍的各种结构和加工过程使我们看到怎样超越构成民间心理学的那种关于信念和愿望的简单化的解释，但一个简单否定意识经验存在的心智理论就像是一个否认热存在的物理学理论，毕竟，我们必须解释为什么人们会有感知特性，这又是怎样发生的。

扩展CRUM

也许一个更为丰富的表征和计算理论能够对说明意识的本质

有所帮助，在本章的前半部分我们已经看到可以把情绪看作一种特别类型的表征，以适合评估、聚焦和行动等功能。其他类型的意识经验也可以同样具有超出了 CRUM 讨论范围内的语言性和图像性表征的表征力和计算力，例如，感知特性就能抓住非常微妙的差别，而这用语言是无法加以刻画的，试想想不同的奶酪或葡萄酒在口味上的差异。切德奶酪肯定在口味上不同于瑞士奶酪或荷兰扁圆干酪，而这一点却很难用语言来描述。葡萄酒品酒师发展了一套复杂的词汇来形容不同的葡萄酒，但这在很大程度上是隐喻式的，无法取代在品尝不同酒类时所体验到的口感和酒香。

意识同样也可以起到将注意力集中到对某一系统来说特别重要的方面去。听到狼在咆哮的经验也许比一个抽象的表征能更有效地把注意力集中到对狼以及逃跑的考虑上去，约翰逊-拉伊德(Johnson-Laird 1983)，提出意识的作用像是大脑的一种操作系统。在当前的计算机里，像 UNIX、DOS、Windows 和 Macintosh System 7 等操作系统所扮演的角色便是按照各种任务所具有的优先权的高低分配处理器资源。同样的，意识所起的作用是根据特定的感知和情绪的重要性来把注意力集中于并行运作的心理过程的某一方面。

意识与计算机操作系统之间的类比在某些方面有一定的弱点，因为与我们所了解的串行计算机的操作系统相比，大脑是一个规模大得多的并行系统，更重要的是，像 UNIX 这样的操作系统在完成任务时，并不需要给计算机提供意识经验。操作系统的比喻提出意识所起的是某种进行中央处理的角色，但意识也可能是由多种非中心式的、并行的、多途径的解释过程集束而成的(Dennett 1991)。扩展 CRUM 以同时从计算过程的原因和效果上来容纳意识经验，要求发展有关表征和计算的新观念。即便是这样，我们仍不能对意识的经验性方面有一个完整的理解。

补充 CRUM

如果我们不打算放弃 CRUM 而接受二元论或某种反计算的观点，我们就需要用生物学方面的见解来补充 CRUM。在对情绪的讨论中，我们已看到当前在大脑结构方面的研究，特别是扁桃核及其与其他结构之间的关系，以及身体其他部位对情绪的作用，对情绪进行生物学解释提供了很好的起点，并可以同思维的计算性解释结合起来。在本章中我从对情绪的讨论着手，是因为目前对情绪的生物学基础的理解要多于对意识的理解。为了从生物学角度研究意识，也许有必要集中于感知的某些特定方面，以理解意识所担当的角色。

克里克（Crick 1994）对视觉意识的神经基础进行了思考。在第六章中我们看到最近对大脑视觉系统的研究对理解视觉表象提供了很大的帮助，克里克试图利用同样类型的神经学发现以求解释视觉意识。从实验心理学出发，他认为意识可能涉及某种形式的注意机制，他推测进行视觉活动的大脑注意某一目标而不是另一目标的这一注意机制涉及相关联的神经元的激励。不同于第七章中介绍的人工神经元，真实的神经元在相互间传播激励是一阵一阵的，这样在同时发生激励的神经元之间可能存在一种协同性，克里克推测神经网络可能具有我们在第七章中所看到的那种竞争性的方面：当某些神经元被激活时，它们会倾向于对其他神经元的激励进行压制。从一个目标到另一个目标之间的视觉注意的转换可能是相互协同的一组神经元联合起来压制另一组相互协同的神经元。

克里克从实验心理学得到的第二个启示是意识涉及短时记忆。用 CRUM 的术语来说，长时记忆是由那些长期存储在记忆中的表征——规则、概念、类比体、表象等等构成的。心理学实验表明短时记忆的容量是非常有限的：如果人们不采用更复杂的结构来进行信息编组，就只能一次记住 7 个条目（Miller 1956）。通

过重复你可以记住你的电话号码，但要记住更长的数字串则需要借助更复杂的编码方式。短时记忆与意识之间的关联在于我们倾向于意识到短时记忆的内容。克里克推测短时记忆的机制可能是相关的神经元具有在一定的时间内激活然后再消退的趋向，或者是通过“震荡回路”使回路中的神经元相互之间保持激励。他介绍了一些实验，在实验中当面对一个视觉目标时，猴子的视神经元发生激励，而当目标移走后这些神经元的激励还能持续上一段时间。如果这些神经元不再激活，猴子在处理这个目标时就易于出现错误，这提示了这些神经元对于完成指定任务的短时记忆来说是十分重要的。

克里克所讨论的注意与短时记忆之间可能的神经关联还不能对视觉经验给出完整的神经学解释。对于注意与记忆的神经学过程以及它们与意识经验之间的关系还有很多东西有待我们进一步了解。但这类研究已为我们展示了这样的可能性，即神经科学有助于将我们的探索集中到与意识有关的各种大脑的活动上。克里克的思路牵涉到大脑的一些部位，比如丘脑，这些部位涉及视觉意识下的注意机制。二元论者可以报怨说无论对意识的神经关联有多深的了解，仍无助于人们想象物质性的大脑如何产生出意识。这种论断有时被称作“基于想象缺乏的论证”。这就相当于说无论对分子运动的了解有多深，也难以想象热只是运动而不是某种像卡路里式的物质。热是由分子运动引起的，对此物理学有足够的依据，与之相比，神经科学对于意识是由神经元激励所引起的证据还远远不够，但认知神经科学的迅速发展会使人们时刻留意对意识的进一步的生物学解释。

放弃 CRUM

如果神经科学的进展能促进对意识的理解，我们能否放弃 CRUM 而指望用神经科学的术语来解释心智的方方面面呢？想想大脑那难以置信的复杂度，有一千亿个神经元和数以万亿个的联

接。直接用神经学术语来说明人类的问题求解、学习和语言的使用着实是不足取的。假定唯物论是正确的，上述的各种活动全都是由神经网络来执行的。但要弄清大脑是如何思维，则要求我们在很多个不同层面上开展工作，虽然克里克和其他人现在已提出了关于视觉意识的神经学基础的理论假说，但目前还看不到直接得出对高层认知能力的神经学解释的曙光。认知神经科学也是一项高度计算化的研究事业，使用简化的人工神经元计算模型来近似地模拟远为复杂得多的大脑的网络。在第十二章中我们会更全面地阐述，增进对心智的理解的最佳策略是将多方面的理论和实验的探索加以综合，包括神经科学、实验心理学和多种多样的计算机建模。要严肃地对待来自意识和情绪的挑战，CRUM 就应当由生物学上的见解加以扩展和补充，而不是放弃 CRUN。

小　　结

情绪和意识在人类思维中的角色需要得到理解，而前面几章中提出的研究路线都不能做到这一点，CRUM 需要加以扩展，以体现针对情绪和意识的更强的表征力和计算力，特别是如何促进鉴别、评估、聚焦和行动等过程的实现。CRUM 还需要由生物学方面的解释加以补充，特别是在对大脑和身体的其他部位的运作怎样促进情绪和意识这一点上，综合式唯物论认为 CRUM 和神经科学可以携手合作，将对高层认知（如问题求解）的解释与经验现象（如意识）的解释结合起来。

讨 论 题

1. CRUM 还可能面临哪些其他的挑战？
2. 情绪对人类思维有作用吗？或着说有助于人类的思维吗？
3. 赋予机器人以情绪需要怎么做？

4. 情绪能看成表征吗?

5. 意识是不是情绪的一个基本性的方面?

6. 在智能思维中意识能起多大的作用?

7. 意识能够从神经元和大脑结构得以理解吗?

进一步的推荐读物

《情绪手册》(Lewis 和 Haviland 1993) 对当前各种不同的有关情绪的心理学研究的状况提供了一个很好的样品，另可参见 Frijda 1986，Goleman 1995，Oatley 1992 以及 Ortony，Clore 和 Collins 1988。

近来关于意识的书有 Baars 1988，Crick 1994，Dennett 1991，Flanagan 1992，Jackendoff 1987 和 Searle 1992。关于对意识的更多的神经心理学上的探讨，请参阅 Churchland 1995、Edelman 1992、Gray (印刷中) 以及 Kosslyn 和 Koenig 1992。

备　　注

我称为综合式唯物论的立场接近于弗拉拉甘 (Flanagan 1992) 所说的“建构式唯物论”和弗斯 (Foss1995) 所说的“方法论唯物论”。保罗·邱奇兰 (Churchland 1989) 和帕特里希·邱奇兰 (Churchland 1986) 坚持排除式唯物论。关于功能主义，见 Johnson-Laird 1983 和 Block 1978。

阿兰·图林提出一种模仿游戏来回答计算机是否能思维。这个游戏后来被称为图林测试。在游戏中，测试者通过电传打字机分别与一个人和一台计算机相联，如果测试者无法分辨出谁是人和谁是计算机，那么我们就可以判定计算机具有智能，这一测试既过松又过严。之所以过松在于一个巧妙编制却几乎不具智能的程序完全有可能糊弄我们一阵子，之所以过严则在于计算机可能在某些无关紧要的人类经验方面有严重不足但在其他方面则体现出很高的智能能力。

创造性经常被列举为对 CRUM 的挑战之一，但前面章节中介绍的几种机制可以用来模拟人类创造性的某些方面，包括逆推、概念组合和类比。另一个有意思的挑战是CRUM、神经科学和 (或) 将在第十一章中讨论的动力

学系统能否解释为什么人会做梦(Flanagan 1995)。布伦勒(Bruner 1990)提出了一种可以称之为**叙述性挑战**,声称对思维的计算和生物学研究忽视了对在故事的解释中人们是怎样相互理解的重要性。

第十章　物质环境与社会环境

本章所讨论的基于物质世界与社会环境两方面的挑战指责CRUM过分局限于心理表征，忽视了思维并不是一个孤立的、脱离具体实现的现象这一基本事实，而强调思维发生在处于复杂的物质与社会环境中的个体身上。这些挑战的极端派全面拒斥心理表征这一观念，主张人的智能是对物质和社会环境的栖居并在其中所进行的操作，全然不同于计算机处理信息的方式。分别针对这两个方面挑战，我将把对认知科学基本假定的多种方式的批评进行汇总，并尝试着给出一个集中的回应。

基于物质世界的挑战

物质世界挑战背后的核心思想是认为思维并不仅仅发生在人的头脑之中，CRUM似乎将思维局限于发生在心智中的计算加工过程，而忽视了人们与物质世界之间进行了连续而丰富的交互作用的事实。这方面的异议有多种不同的形式，从受到马丁·海德格尔(Martin Heidegger)影响的哲学家到对人工智能中正统路线不满的机器人学的研究学者。

存在于世

人们是怎样用锤子钉钉子的？如果从CRUM的角度来回答这个问题，我们首先要考虑对于锤子和钉子采用何种表征。我们可能使用概念或表象来代表锤子和钉子，而钉钉子这个行为之所以能发生是因为我们能够在这些表征上进行计算操作，并最终转换为用锤子钉钉子这一物理运动。

德国哲学家海德格尔拒斥这样一种有关钉钉子的表征观(Heidegger 1962; Dreyfus 1991)，他否认认知科学所设定的在表征物与外部世界的分离，并主张我们在这一物质世界中发生作用仅仅是因为我们就是这个世界的一部分，他用“存在于世”(Being-in-the World)这一表述来表明人们能完成钉钉子这样的任务是依靠人们自身的物理技能，而毋须任何类型的表征。德雷弗斯(Dreyfus 1992)提出了海德格尔式的论证来反对人工智能，宣称其试图进行形式化表征知识的做法是没有希望的，因为我们的智能是内在地非表征性的。

反表征的认知观在一些不满传统人工智能路线的研究者中受到欢迎，温诺格拉德和弗洛尔斯(Winograd 和 Flores 1986)赞成海德格尔式的观点并得出了人工智能的正统路线是不可能成功的结论，因为我们不可能表征作为人类各种能力基础的大量的背景信息。史密斯(Smith 1991)提出他所谓的“嵌入式计算”，通过侧重于与物质世界的交互作用而不是内部的加工处理，来避免传统计算方式的表征负担。

机器人学

与之相似，布鲁克斯(Brooks 1991)提出了一种与基于逻辑和规则的传统大异其趣的建造机器人的途径。不同于将一台高层次机器人在环境中对其运动进行规划所需的知识进行编码，他建造的是具有从环境中进行学习的能力的简单机制的机器人。不是将关于如何行走的复杂规则进行编码，布鲁克斯给他的类似昆虫的机器人装上多个处理器，使它们在与环境的交互中学习行走。其中没有使用前面讨论过的六种主要的CRUM方式的任何表征技术。马克沃斯(Mackworth 1993)也提出将机器人建造成嵌入在真实世界中的物理系统，将知觉与行为紧密地耦合起来。他建造了一个移动机器人系统，能敏捷地进行一种简单的足球赛，其中没有使用基于传统人工智能的机器人通常要用到的复杂的表征和

规划。

情境化的行动

一些人类学家和心理学家也提出认知科学过于强调心理过程的角色而低估了情境和背景在人的问题求解和学习过程中的角色。苏彻曼（Suchman 1987）及雷弗和温格（Lave 和 Wenger 1991）反对认知心理学过于依靠人工的任务来研究人的思维，真实环境中的问题求解并不是那么依重于心理表征，而是借助于与外部世界和其他人的直接交互。缺少抽象数学表征的人也可以将一些特定的任务做得很出色，比如在晚会上分配一块比萨饼。使用一台复杂的机器例如计算机并不需要对它形成抽象的表征，而是学习与它进行交往。苏彻曼和雷弗把人看作是通过与世界的交往来进行思考的，而不是对世界进行表征并处理这些表征，这就像布鲁克斯的简单机器人一样。夏农（Shanon 1993）指出 CRUM 无法说明人们与世界发生关系的那些微妙的、依重具体情境的方式。

躯体与直接知觉

人们是怎样与世界进行交互的呢？来自外部世界的信息需要通过感知进行转换进入心智，CRUM 将知觉视为对反映外部世界特征的表征的推理或建构。受吉布森（Gibson 1979）影响的心理学家反对这样一种知觉的推理观，他们认为我们能更为直接地了解这个世界。我们的知觉器官能与外部世界相协调，这样信息就直接输送到大脑而不需要在表征层面上进行计算，我们的生理感知器官参与构成我们与外部世界进行交互的能力。

正如我们在对情绪的讨论中所看到的那样，CRUM 倾向于在本质上将我们的躯体视为与我们的认知活动基本上不相关。但约翰逊（Johnson 1987）和拉可夫（Lakoff 1987）却指出这一传统忽视了我们的躯体在我们的思维活动中所扮演的关键性角色。散布

在我们语言中的许多隐喻实际上来自于以身体为基础的关系，比如上与下、左和右以及内与外。如果我们不是拥有我们这样的躯体，不是活动在我们寄居的世界中，那么我们的隐喻系统和我们的整个心理装置将会大不一样。CRUM 的一大优点是有可能将其心智观应用于计算机和外星生命，而不依赖其特定的物理构成，这看上去似乎是颇为诱人，但如果人类思维的关键之处依赖于我们所特有的这种躯体以及它如何与外部世界相协调，那么这一优点就是虚幻的。

意向性

心智与世界的关系中还涉及到像约翰·塞尔(Searle 1992)等哲学家视为对 CRUM 的一个严峻挑战。心理状态是用以表征这个世界的：它们具有意向性，它们是**关于**某种事物的。你认为你的朋友正在图书馆这么一个信念不只是你头脑里的一个表征，它是关于你的朋友和图书馆的，而它们是这个世界的一部分。塞尔用了一个思想实验来表明计算机不可能具有意向性，请想象你被锁在一间房子里，有人从门外递给你一些纸张，上面有许多你不认识的符号。然而，你却能够利用一套指令表查出这些符号，并且挑选一些其他符号然后递出房间。这些符号是中文，这是你所不认识的，而当你传出那些你查出来的符号时，你对你收到的问题给出了有意义的答案。塞尔指出很显然你只不过是在操作你并不理解的符号，同样，计算机操作符号时也与理解无关。人们操作的符号具有意向性这样的语义性质，这是基于我们与世界的交互，而计算机里的表征是独立于外部世界的，因而缺少意向性。计算机是纯粹的句法机械，缺乏人类的语义能力，不能够在与世界交互作用的基础上赋予符号以意义。所以，心智的计算观在根本上是错误的。

对基于物质世界挑战的回应

否　认

认知科学可以忽视外在物质世界吗？尽管CRUM注重表征和过程而不是物质性的交互，CRUM不应当简单地否认基于外在世界的挑战。心理学和人工智能的研究者有时用“表征”一词来指示内在的“结构”，但结构成其为表征只有当它是用来代表某件其他的事物。

在人们解决问题和学习的过程中，人们并不像与世界脱节分离的计算机那么运作，至今为止所发展的计算模型都倾向于忽略物质环境的细节，用于建模的大多数计算机除了编程者用来输入指令的键盘以外就没有与这个世界的联系。对外在世界结构的重视对于设计人们更易于使用的机器和工具是十分重要的(Norman 1989)。因此，CRUM需要得以扩展和补充以容纳躯体和外部世界。存在于世、嵌入式计算、情境化行动和意向性等论点的激进派会把这种扩展视为在浪费时间，因为CRUM完全走入了歧途，但CRUM在解释上的成就使之有资格去面对基于外在世界的挑战。海德格尔也许会说我们并不是在表征这个世界，我们只是体现它，成为其中的一部分，但这也许不是唯一的选择。

扩展CRUM

正如我们在第九章里所看到的，情绪和意识提示我们需要扩展表征的范围，不应局限于在第二章至第七章里所讨论过的标准的CRUM表征。同样，重视外在世界与躯体要求发展新的表征类型，当我看见山顶上的一所房子，语言编码**在之上**(**房子，山**)并没有抓住我的感知经验所告诉我的所有的知识。表象式表征不光包括在第六章讨论过的视觉表象，还包括从其他感知器官如嗅、味、听和触觉得来的表象。用锤子钉钉子不仅仅是表达锤子和钉

子的概念，它还是对手臂挥动锤子的运动感觉的表征，不过，借助从物理性和神经学方面的抽象，对 CRUM 的这种扩展对于给予 CRUM 解释与世界进行交互作用的能力而言只能说是一条羊肠小道。

补充 CRUM

正如大脑和躯体的结构与运作对理解情绪有关，对与世界进行交互的全面的说明也应涉及生物学方面的内容。科斯林（Kosslyn 1994）在表象方面的工作为我们提供了一种可能的模式，视觉表征与大脑如何对它们进行加工是密切相关的。同样，我们可以指望通过了解大脑如何处理各种感知输入，来推测他们与我们的言语表征之间的关联。并不是所有的大脑加工过程都必须视为计算：参见第十一章对大脑作为动力学系统的讨论。但是对人类思维的完整理解必须将大脑的非表征运作与计算程序结合起来，后者对于高层认知是至关重要的。如果说某些 CRUM 的支持者忽视了我们能行走、投掷和挥锤这类事实，那么极端的环境论者的不足则是要么把人类的思维消解为简单化的、昆虫似的反应，要么将其视为神秘的、无法分析的偶然事件。

放弃 CRUM

然而如果在我们的图景中加入了神经科学，为什么我们还需要 CRUM？在外部世界与大脑的神经网络之间有这么一个不必要的解释的中间层，为什么不把它从我们的图景中拿掉呢？答案就在于人类问题求解的多样性和解释上的复杂性，布鲁克斯的类似昆虫的机器人能在其所处的环境中灵活自如地学习行走，这一能力给人留下很深刻的印象，但它们完全缺乏人所具有的那种安排到达另一个城市的高层的规划能力。同样，情境化对于简单性的任务而言是一大特长，但人能解决复杂的、抽象的问题，远远超出了对环境的简单响应。表象、概念、规则和其他的表征使我们

能对外部世界进行想象化的操作，而不依赖于实际的操作行为。

意向性是一个严峻的问题，但它并没有超出经过适当增强的CRUM 的适用范围。我们必须承认目前用于模拟思维的台式计算机是缺乏语义学性质的句法装置：在我的计算机程序里写了**啤酒**二字并不意味着计算机能理解啤酒是什么。然而通过给计算提供对外部世界的视觉、听觉和触觉输入的机器人界面，并且配上能生成各种分布式的、语言的或图像式的表征的学习算法，计算机已经是具备了从外部世界进行学习的有限的能力。塞尔的“中文房间”的例子通过使你把自己视为一个与外界隔离的符号处理器，调动了你的直觉，使你认识到了纯符号加工的局限。而你之所以能有效地处理中文符号是因为你是整个系统的一个部分，整个系统与外部世界发生交互作用，由此掌握了对照表上的符号，这样你加上你所利用了的整个装置，就能自然地判断为是理解中文的。同样地，一台机器人式的、能学习的计算机能在处理句法时同时解决语义，因而对它的内部表征赋予了意向性。为了对人类的思维提供一个完整的模型，CRUM 必须更为重视外部世界以及我们与之打交道的躯体，这样的扩充是可能的，也是自然的。

基于社会环境的挑战

正如认知科学忽视了思维者所处的物质环境，它也倾向于忽视思维者所处的社会环境。哲学、认知心理学和人工智能主要关注的是发生在个体身上的心理表征与过程。不过，这些领域最新的发展动向已转向注重知识的社会背景，基于社会环境对CRUM 的挑战关心的是其能否加以扩展或补充以处理思维的社会性方面的问题。

社会认识论

认识论是哲学的一个分支，关注知识的本质以及知识的辩护。

传统的认识论关心的是个体所知道的是什么：如何辩护我们的信念？不过，知识的社会性方面的问题越来越多地受到重视，特别是在科学哲学家当中。对现代科学来说，知识显然是一项社会性的事业，大多数发表的论文都有不只一位的作者，很多的研究也是有一组共同工作的科学家合作完成的，最近发布的确定“顶夸克”的工作是由400多位物理学家联名发表的。许多其他类型的知识也是内在的社会性的，从公司中合作工作的同事共享的商业知识到交响乐和戏剧中的艺术知识。

分布式认知

心理学家和人类学家也对知识的社会性方面表现出越来越大的兴趣，特别是对一些应用领域，如教育与工作场合。认知被视为是“分布式的”，不是在单个个体的心智中发生，而是通过多个个体的相互合作。例如，一组软件工程师合作开发一个新的计算机软件就要确保能达到一个共同的目标。他们必须形成对该软件所要完成的任务的一个共同的表征，他们还需要保持不断的交流以确保整个项目的不同部分能够合起来工作。类似的交流与合作对于完成任何一个多人参与的任务都是必不可少的。在第二章至第八章中评价的各种CRUM途径所讨论的问题求解都是针对单个个体的，但在当今世界对问题的解决通常都是由群体来完成的，学生有时完成课程作业时要按分组来进行，任何一个参与过如编辑报纸或俱乐部工作等课余活动的人都很清楚与他人合作完成共同的目标是何等重要。

分布式人工智能

在其发展的最初几十年中，人工智能关心的是怎样让单个的计算机完成智能任务。不过，近来越来越多的注意力放到了如何让计算机网络共同工作。大学校园里有数百台计算机通过网络互连，而Internet在全世界连接了成千上万台计算机。不同于在单个

计算机里建造一个完整的智能系统，有可能在多台计算机中分布多种不同侧重的专门智能能力。通过相互交流，多台计算机能够解决一台计算机单独工作所不能解决的问题。分布式人工智能(DAI)是人工智能中的一个新的分支，研究拥有不同类型知识库的计算机怎样连接并进行合作。试想有四台不同的计算机，分别运行基于逻辑、规则、类比和分布式表征的专家系统，它们能够共同工作，克服交流上的困难，并且产生出胜过单个系统能力的专家系统吗？

分布式人工智能有些类似于联接主义（第七章），都涉及多个处理器的并列加工，所不同的是一个分布式人工智能系统中的每一个处理器本身都是一个拥有某些智能的高级系统，相反，联接主义系统中的单元则非常简单，只能向其他单元传递激励，而不能像分布式人工智能系统中的计算机那样传递复杂的消息。

文化与知识的社会建构

不同于哲学家、心理学家和人工智能研究者，人类学家一直关注着认知的社会性方面。人类学的中心概念是文化，“这是约定俗成的意义系统，作为任何特定社会中的行为规范，或行为准则”(Barrett 1991，第55页)。文化的概念显然是社会性的，因为它关心的是那些使得社会交往成为可能的、整个社会共有的信念和价值。作为对社会背景中的认知的研究的一部分，一些心理学家近来也将注意力转到了文化上，赫希菲尔德和格尔曼(Hirschfeld 和 Gelman 1994，第4页）主张心智“与其说是一个通用目的的问题求解者，不如说是针对各种与环境相关任务的、稳定且独立的子系统的集合”。这些任务可能与稳定的文化相关，因而问题求解会呈现出文化上的多样性。同样，情绪可能不能完全置入人类的思维中，但会随文化的不同而不同（Kitayama 和 Markus 1994)。

一些社会学家过于看重文化多样性的思想，而拒斥任何客观

性知识的观念。对他们来说，整个世界是“社会建构性的”，人们的信念系统不是来源于他们的认识过程，而是来自他们的社会环境。对知识的认知上的解释的怀疑论在科学的社会学中很常见，例如 Latour 和 Woolgar 1986。

对社会性挑战的回应

否 认

各个方面的论证要求思维不应当理解为发生在社会的真空中，而应视为内在的交互式的过程，CRUM 该怎样回应这样的挑战呢？对此的否认可以采取社会科学哲学家所称的**方法论个体主义**的形式。这种观点认为群体行为只不过是个体行为的聚合，在解释群体行为时没有必要超出对个体行为的理解。方法论个体主义在经济学家中很流行，对国家、团体和社会阶层等宏观经济群体的理解原则上都可以由对个体行为的解释来取代。同样，对个体性的 CRUM 的辩护也可以坚持认为来自群体、网络和文化等社会性方面的挑战可以逐步通过对个体的理解来消解。

我们可以承认群体、网络和社会是由个体组成这一物质事实，而同时强调对它们的解释不能忽视社会结构。社会是异常复杂的系统，它们的运作是动态的和交互式的，通过还原到对个体的解释是很难获得对系统的理解。在第九章中我们也看到了同样的复杂性问题，大脑的运作也许是由神经元的活动组成的，但数十亿个神经元以高度复杂的方式相互作用，从神经元行为的基础上看来不大可能完整地说明个人层面的行为，更有希望的方式是重视各个层面上的解释（神经元、个人、社会），并且研究这些层面相互间如何发生联系（参见第十二章）。

扩展 CRUM

重视社会性方面的挑战要求 CRUM 从某种不同的角度来审

视表征和加工过程，心理表征的作用不仅仅由个体来完成，也不仅仅为个体所利用，还为群体所共享和使用。命题、规则、概念、表象、类比甚至分布式表征都需要从一个个体传递到另一个，例如：类比就不只是一个人单独去解决问题的过程，它还可以是一个人帮助另一个人解决问题的重要方式，由此形成了对某一情景的共享式的表征。当你要解决如何选课或如何获得一个计算机帐号这样的问题时，你很可能要靠别人提供信息。

由于表征要具有社会性方面的可应用性，就必须有个人之间的操作过程来促使表征能从一个人传递到另一个人。在计算机网络上这样的传递乍看起来似乎很简单，因为电子链路使得电子邮件的传输看上去似乎轻而易举，其实，看上去简单的传递要依赖于协议的建立，这样各种不同硬件和软件的计算机才能够相互通讯。这样看来，分布式人工智能就远不是那么平淡无奇。同样，人与人之间的交流通常也是非常困难的，教学不仅仅是要把信息填进学生的脑袋里，而是要把特定的表征系统传给学生。因此，我们要扩展 CRUM，以包括对表征在个体之间传递过程的描述。

补充 CRUM

对认知的社会性过程的研究，包括心理学的、计算的、认识论的和文化的诸方面，才刚刚开了个头。尽管 CRUM 可以在社会性方向上进行扩展以包含有关表征和计算的更强的观念，我们仍然应当指望对思维的理解还需要借助于一些内在的社会性概念，如群体、网络、社会、文化和交流。来自情绪、意识和外部世界方面的挑战已表明需要用生物学方面的见解对 CRUM 加以补充，同样 CRUM 也需要加上社会性方面的考虑。我们可以把这种研究心智的理想化途径称为 CRUMBS，即**生物-社会性的，对心智的表征-计算理解** (Computational-Representational Understanding of Mind, Biological Social)。这一缩写所代表的是认知科学作为一项整体性事业的分崩离析，取而代之的是从各个局部出发对

心智的多重理解。但是进展将从CRUMBS的各个不同方面推进，而目前尚没有更好的途径使我们对整个人类精神现象得到更好的理解。

放弃 CRUM

如同极端派的海德格尔主义者和情景化行为的信奉者，一些社会结构论者主张全面放弃CRUM，而采用关于知识的纯粹的社会性解释。虽说人类思维的社会性难度的重要性不容忽视，我们也不应该忘记问题求解、学习和语言同样也可以由个体心智的表征与过程来加以理解。心智与社会是互补性的解释观念，而不是竞争者，因此，对认知科学的社会挑战应视为对扩展与补充CRUM的激励，而不是要抛弃它。综合式唯物主义，这一在第九章中提出的立场，能够引起我们对社会性和生物学方面的重视。

小　　结

从来自外部世界的挑战看，认知科学对于心智与它们所处的物质环境之间的关系没有引起足够的重视，这一挑战既来自抽象的形式（意向性问题，涉及表征如何成为关于外部世界的东西），也以具体的心理学和计算上的形式出现。某些CRUM的批评者认为我们的心智不需要去表征外部世界，因为它们就置身于这一世界之中。虽然CRUM需要加以扩展和补充以便更好地描述思维是如何依赖于与外部世界的交互的，但过于强调环境而排除了表征的心智观却不足以全面说明人类的智能行为。同样的，社会性方面的挑战并没有构成对CRUM的取代，而是指出了群体、网络、社会和文化等认知科学需要加以重视的问题，CRUM有必要进行扩展和补充。

讨论题

1. 驾驶汽车的过程中有哪些部分要涉及心理表征，而有哪些只要求知道如何与世界进行交互？

2. 计算机可以具有意向性吗？

3. 心智的计算机模型必然要忽视行为的物质背景吗？

4. 对情绪的理解怎样有助于研究分布式认知？

5. 你指望一个计算机网络比一台计算机单独工作具有更多的智能吗？

6. CRUMBS 是一个融贯一致的观点吗？关于心智的理论可以既是生物性的和社会性的，同时又是计算性的吗？

进一步的推荐读物

《认知科学》杂志在 1993 年元月曾就情境化行为出过一期专刊，包括有激进的支持和反对两方面的意见。Dietrich 1994 收录了捍卫心智的计算观，反对包括意向性在内的各种反对意见的论文。

在社会认识论方面近期的著作有 Giere 1988，Goldman 1992，Kitcher 1993，Schmitt 1994，Solomon 1994 及 Thagard 1993，1994。关于分布式认知，见 Galegher，Kraut 和 Egido 1990，Hutchins 1995，Resnick，Levine 和 Behrend 1991，及 Salomon 1993。关于分布式人工智能，见 Bond 和 Gasser 1988，Durfee 1992，Durfee，Lesser 和 Corkhill 1989，Gasser1991，及 Hewitt 1991。

备　　注

关于表征是如何进行表征的这一问题存在相当多的哲学讨论。其中一个问题涉及外部主义（externalism），强调表征的内容是由外部世界决定的。Von Eckardt 1993 包含了最近有关内容确定问题各种论点的一个很好的探

讨；另可参见 Fodor 1987。在第四章中我论证了概念的意义既与它们之间的相互关系有关，又涉及它们与外部世界之间的联系。

动力学系统（第十一章）也提出外部世界方面的一种可能的挑战。

第十一章　动力学系统与数学知识

动力学系统的挑战

在第一章的小结中我们给出了基于对心智的表征-计算理解的解释程式，并在第二章至第七章的小结里用各种特定的表征与过程给出该程式的例示。有许多的科学解释并不使用这种解释模式。假定我们要解释为什么昨天下雨了，没有一个严肃的气象学家会说云层和雨滴有什么样的信念或目的，这些东西就没有心理表征。相反，气象学家会采用一大堆的变量，包括各种场合的温度、湿度和气压，来作出他们的预见和解释。他们把这些变量代入到用来描述气象系统如何随时间发生变化的数学方程中。气象学家把气象称为**动力学系统**，也就是说，看作随时间发生的变化可以用一组方程式来加以刻画的系统，变量的当前值在数学上依赖于那些变量先前的值。

物理学、生物学乃至经济学中的许多现象都可以用动力学系统的概念来理解，如状态空间、吸引子、相变和混沌。一个系统的**状态空间**是指可以由测量该系统的变量来确定的状态集，例如，一个十分简单的气象系统模型可以测定 5 个不同地点的温度、湿度和气压，共有 15 个变量，这些变量值的所有不同组合就构成了该系统的状态空间。该系统的变化可以由在这个空间中的一个点（所有变量值的一个组合）到另一点的运动来加以刻画。在高速计算机出现以前，科学家们解决其变化只能用线性方程来描述的简单系统。线性方程的形式如

$$y = kx + c.$$

在这个方程式中，变量 y 的值只取决于变量 x 的值乘以常数 k 再加上一个常数 c。但复杂动力学系统必须由非线性方程来描述，如 $y = xz$，这里变量 y 的值依赖 x 与 z 的相互关系。

非线性系统具有非常奇特的行为，会在短时间内从状态空间的一个点跳到另一个非常不同的点。例如，随着冷空气前锋的到来天气会在一两个小时内发生剧烈变化，风雨大作。尽管有这些剧烈的变化，动力学系统也会有趋向于常居的、相对稳定的状态，称为**吸引子**。一个系统可以有多个吸引子，所以可能有不止一个的稳定状态。从一个吸引子状态到另一个吸引子状态的变化构成了一个**相变**，比如天气从寒冷干燥转变为湿热，或者是水冷到一定程度会发生凝固。在这两种情况里，看似微小局部变化都使系统转入了性质上非常不同的状态。

如果一个动力学系统对初始条件极为敏感，该系统就会表现出**混沌**，这也就是说，如果方程式中变量的值发生微小的变化，随着系统的演变会产生出完全不同的结果。天气就是一个混沌系统，因为在很远地方的一些微小的气象变化经过时间的累加会引起本地天气的剧烈变化，这被称作"**蝴蝶效应**"：在中国的一只蝴蝶扇动它的翅膀在当地气象系统中引起的微小变化可能会导致别的地方天气的剧烈变化。由于对诸多变量的细微变化极为敏感，混沌系统能呈现出很难预测的剧变（相变）。很难提前几天对天气进行预报的原因之一就在于气象学家无法对影响未来几天天气变化的所有变量的所有细微差别进行测量。

动力学系统对认知科学构成的挑战是：我们应当把心智视为一个动力学系统，而不应将人类思维理解为表征-计算的形式，不同于提出一整套的表征与计算过程，我们应当像在物理学和生物学中成功的那样用方程式来刻画心智是如何随时间而发生变化的。下面是反映了某些动力学系统观念的解释程式：

解释目标

为什么人们会具有**稳定**的但**无法**预见的行为**模式**?

解释模式

人类的思维由一组**变量**来刻画。

这些**变量**服从于一组非线性**方程**。

这些**方程**建立了一个具有**吸引子**的**状态空间**。

由这些**方程**所刻画的这一系统是**混沌**的。

吸引子的存在解释了行为的**稳定模式**。

多重吸引子解释剧烈的**相变**。

系统的**混沌**本质解释为什么行为是**不可预测**的。

动力学系统挑战的合理性取决于这样的解释程式能从多少方面说明人类思维的程度。

将动力学系统的概念运用于认知只是发生在最近：大多数文献发表于90年代。可以有三种不同的方式将心智视为动力学系统。像在物理学和生物学中如此强有力的解释模式直接能运用于认知的情况是比较少见的。在心理学里很难确定少量的相关变量并用它们写成方程式来进行有益的预测。尽管如此，使用少量变量和方程的动力学系统模型已用于决策制定（Busemeyer 和 Townsend 1993；Richards 1990）和语言发展（Van Geert 1991）等认知现象。

较为普遍的情况是，研究者隐喻式地使用动力学系统概念，尚无法确定相应的变量和方程。即便这些东西还不甚清楚，用状态空间的变化、吸引子、相变和混沌等来描述复杂系统的变化仍是可能的。特伦和史密斯（Thelen 和 Smith 1994）用吸引子状态的变迁来解释儿童学习行走。另外，在临床心理学里也隐喻式地使用了动力学系统的概念（Barton 1994；Schmid 1991）。

动力学系统概念的第三种应用被部分联接主义者所采纳。他们发现用动力学系统的术语来描述人工和真实的神经网络极为有

用。联接主义系统是明显的动力学系统，包含代表各种单元激励值和单元联接强度的变量，以及修正这些激励值和改变联接强度的非线性方程。与上面提到的第一种情况不同，联接主义动力学系统中的变量数目非常大，因为对每一个单元都有一个变量来代表它的激励值。像吸引子、相变和混沌等概念都可以用于神经网络：例如，当一个网络安定下来时，就要求有一个稳定的状态，不再发生激励值的变化。用动力学系统的概念来刻画的联接主义模型有波拉克（Pollack 1991）及斯卡达和弗里曼（Skarda 和 Freeman 1987）开发的系统。

对动力学系统挑战的回应

否　认

对 CRUM 的维护可以指责动力学系统的方法在对人类思维研究的应用中是十分有限的。虽然这种方法在物理学和生物学中十分有效，但确定少量的方程和变量在心理学研究中就远不是那么有用。将动力学系统的概念笼统地加以应用的情况虽不少见，却很难进行精确的预测和建模。联接主义模型是比较精确，但联接主义却是 CRUM 的一部分，而不是它的对手。所以动力学系统最好是视为联接主义的附庸，而不是 CRUM 的敌手。动力学系统思想在思维研究中的上述三种方式的运用都暴露了其局限性：人类思维不适合于用少数变量来描述，笼统的解释没有太大的用处，而联接主义顶多也只是动力学思想的微乎其微的应用。

扩展和补充 CRUM

动力学系统的途径应该受到更多的重视，而不只是像上述论证所提议的那样，因为它能证明 CRUM 中相对被忽视的几个重要方面。首先，正如范·戈尔德和波特（Van Gelder 和 Port 1995）所指出的那样，动力学系统比 CRUM 能更好地处理时间，

它提供了一系列新的概念来描述在智能系统中所发生的变化。其次，动力学系统为避免第十章中所讨论的来自外部世界的挑战提供了一种可能的途径。我们知道CRUM面临的问题是，心智不是与世隔绝的处理器，而必须与变动不居的外部世界发生交互作用。从动力学系统的观点看，心智与世界不是相互分离的，而是共同构成了一个大的动力学系统。要用方程式来描述心智与世界的相互作用显然是太难了，但至少心智与世界是用一组相匹配的术语来处理的。第三，动力学系统可能对解释人类行为的一些非表征性的方面会很有作用。即使问题求解和语言最好是用心理表征来理解，人类行为的其他方面，如运动控制、情绪和睡眠，用动力学系统的概念来解释就来得更为自然。当一个两岁的儿童因为一些微不足道的小事顷刻间从开怀大笑变为嚎啕大哭，行为上的这一剧烈变化所涉及到的东西要比对规则、概念、类比、表象或分布式表征的加工要多得多。情绪上的变化，以及儿童的蹒跚学步或一个人渐入梦乡，用动力学系统的吸引子和相变来理解可能会更合适一些。

因而，CRUM，特别是联接主义这一支，对用动力学系统思想来加以扩展和补充看来是开放的。心智是一个动力学系统这一假说也不是CRUM的一个够格的对手，因为有很多关于问题求解、学习和语言的现象是可以用CRUM来解释的，而这些是动力学系统的鼓吹者们还未曾谈到过的。不过，容纳了人类生物学和与世界交互作用的对心智的全面地说明应该能从动力学系统的解释中获得有益的启示。

数学上的挑战

在第二章我们考察了形式逻辑作为理解心智的一种计算-表征的方式，对于像古特罗伯·弗雷格和伯特兰·罗素这样的数学家来说，发展现代形式逻辑的很大一部分的动机是要为数学知识

提供一个说明，他们试图用逻辑从少量的基本假设推导出所有的数学知识。然而，在 30 年代，奥地利数学家库尔特·哥德尔证明了这样的计划是不可能实现的，他证明了一条著名的定理，表明没有一个形式化的系统能够刻画所有的数学知识。哥德尔不完备定理指出任何一个一致性的并且能够对算术进行形式化的系统都是不完备的，也就是说该系统中会有一个公式既不能被证明也不能被否证。哥德尔向人们展示了如何为每一个满足一致性并且足够强的形式化系统构造一条公式，该公式在这一形式化的系统中既不能被证明，也不能被否证。令人吃惊的是，对所构造的这一公式的最自然的解释是它本身是不可证明的。因而，这条公式是一条正确的算术表达式，但它不能在这一形式化系统中得到证明，所以说这一形式化系统对算术来说是不完备的，因为有一条正确的算术表达式在这个系统无法得到证明。

一些哲学家借助于哥德尔定理来说明对心智的计算解释是不可能的（Nagel 和 Newman 1958；Lucas 1961）。下面是这一论证：

1. 任何一个试图作为人类心智模型的计算机都是一个形式化系统的实例。

2. 如果这个形式化系统是一致性的并且适合于说明算术，那么根据哥德尔定理它就是不完备的，至少有一条公式既不能得到证明也不能被否证。

3. 但是人的心智可以知道这条公式是正确的，所以有些事情是心智可以做而计算机不能做的。

4. 因此，心智不是计算机。

我们由此得到了一个数学上的论证，至少针对数学知识来说，人的心智超过了任何的计算机。如果这一论证能成立，那么其结论就是通过计算的方式去理解数学知识是有局限的。然而，数学哲学家对这一论证中涉及哪些事情是人能做而哪些事情形式化系统

不能做的假设提出了置疑（Benacerraf 1967）。

在罗杰·彭罗斯的两本书中（Penrose 1989，1994）这种反计算的论证又以一种更为激进的方式再现。彭罗斯提出了他自己的哥德尔定理版本用以论证数学知识不能从计算上进行理解。他认为人类的心智具有一种领悟无可辩驳的数学真理的能力，也就是说能看出某些命题在数学上是无可怀疑的。但是，他不是一个二元论者，他将人类心智所具有的数学能力视为神经元范围内的量子效应所导致的特殊的非计算性质。因而，心智是一台量子计算机，具备目前为人类思维提供类比的各种类型的计算机所不具有的特殊性质。

彭罗斯反对心智的计算观的哥德尔式的论证借用了图灵机的概念，这是一种非常简单的计算机，含有有限的内部状态集和一条无限长的纸带，纸带上划分为方格，要么标上 1，要么标上 0。图灵机所要做的是从纸带上读出标记，参照其内部状态的设定，然后改变其内部状态，并在纸带上写下新的标记，最后停机。令人惊奇的是，可以证明数字计算机可以做的任何事情，图灵机都可以做，所以图灵机对于什么是可以计算的提供了一个普遍性的概括。在现代计算机上可以运行的任何算法在原则上都可以在图灵机上运行。但有些事是图灵机所无法做的，例如确定它自己是否会停机。

彭罗斯（Penrose 1994）表明对任何图灵机来说都有一种计算是人类的数学家可以知道是不会结束的，即便机器本身不能证明该计算不会停止。他把这一结论称为 G（代表哥德尔），其中一个重要的限定是计算是可知为可靠的。

G 人类的数学家在确认数学真理时不使用某种可知为可靠的算法。

彭罗斯的假定是任何可以确认数学真理的计算机都要使用某种可

知为可靠的算法，因而，没有任何计算机能够模拟人类的数学知识。在这里，“可靠的”指的是“不会给出错误的答案”，而“可知地”意味着我们能够知道该算法是可靠的。概括为要点，彭罗斯的论证如下：

1. 计算机可以做的任何事情，图灵机都可以做。所以图灵机所不能做的任何事情，计算机也不能做。

2. 对任何图灵机 TM，我们都可以设计一件它不能完成的任务，即定义一种计算，TM 不能确定它是否会停止。

3. 但数学家如果知道这种计算是可靠的（一致的，不会有错），他们能够确定 TM 不会停机。

4. 所以人类可以做 TM 所不能做的某些事情。也就是说，计算机不能辨认数学真理。

彭罗斯是否已终结性地表明数学知识是非计算的呢？

对数学上的挑战的回应

否　认

有关数学知识的可靠性、可知性和一致性的一些问题表明我们可以接受 G 和哥德尔定理而毋须拒斥数学知识的计算观。让我们以一种比图灵机具体的方式，通过谈论复杂的表征和加工过程来展开我们的探讨。为了更真切地模拟人类数学家，我们也许需要在第二章至第七章讨论过的**所有**类型的表征。数学家们显然使用了诸如“所有的”、“有一些”、“如果-那么”、“与”和“或”等关系的操作，所以我们需要形式逻辑为我们提供的所有表征资源。此外，计算系统会需要某种类似于形式逻辑的演绎能力的东西来模拟数学家如何进行数学证明。也许基于规则的系统用于构造定理证明会远比用形式逻辑编制的定理证明器来得更灵活些。我们

还需要丰富的资源来表达数学概念，比如**数**和**可除尽**，还包括用于产生一些重要的概念组合如质数的程序。类比在数学知识的发展中也扮演着重要角色，数学家们经常从熟悉的数学领域中借用一些概念和技巧去发展新的领域（Polya 1957）。表象在数学思维中也很重要，特别是在一些与视觉有关的分支，如几何学和拓扑学。构造一个图表或心理图像可能对构造一个证明及判定某一特定命题会提供很大的帮助。

联接主义的思想对理解数学知识的本质可能也有相关之处，我们可以将概念理解为分布式的表征，我们可以将对一组公理的采纳视为一种并行的约束满足。在数学上通常并不只是选定一组无可置疑公理，然后看看能从中推出些什么东西来。有时候，公理之所以被采纳是因为从中能导出有意思的定理。对公理集和定理的选择就要依赖它们之间的相互关系，而这便可以置入并行约束满足的模型中。

很显然，建造一个全面完整的人类数学家的认知模型是一项巨大的工程，远远超出了目前的心理学和计算上的知识。不过，让我们想象未来的认知科学家能建造出一个称为CAM（**认知算术模型**，Cognitive Arithmetic Model），将各种类型的表征和加工过程包容在内，那么哥德尔定理或者彭罗斯的图灵机论证还能用来说明CAM从根本上就赶不上人类数学家吗？

要使彭罗斯的论证成立，我们要能够构造出一台与CAM等价的图灵机。计算理论告诉我们这样一台图灵机是存在的，但这不保证任何人类数学家能够构造出它来。用以实现CAM的程序更像是一大群程序员合作工作的结果，没有一个人会知道整个系统是如何动作的。很可能是一位程序员建造一个类比机制，另一个则负责定理证明器，如此不一而足。就像当今产业界生产的其他大型软件工程，CAM的代码足有百万行，而单个程序员的知识仅限制于少量的模块。由于没人能了解CAM所包含的整个算法集，没人能够将CAM转换为一个图灵机。对于习惯了用高级编程

语言的人来说，在图灵机上完成哪怕是一个简单任务的编程都是令人痛苦万分的。将 CAM 转换为图灵机也有可能令其自动进行（参见下面的备注部分），但没有人能够完全理解这一转换的结果，它比 CAM 的输入代码复杂得多得多（其动行也会慢得多得多）。

因此没有理由认为任何一个个人能够构造出与 CAM 等价的图灵机。对人类认知的理解，包括发展和鉴别数学知识的能力，不是来自于借助抽象计算的思维，而是根据解释数学思维的强有力的表征与加工过程。

所以，用以实现 CAM 的图灵机算法或形式化系统可能对于任何一个数学家来说是无法了解的，虽说不同类型的高层算法是由建造 CAM 的不同的成员所掌握。即使某一位数学家能够发现图灵机与 CAM 等价，但没有理由指望有可能去证明 CAM 是可靠的或一致的。彭罗斯否认数学家们使用了不可靠的方法，虽然他承认他们有时也会出错。例如，安德鲁·威尔斯在 1993 年给出了对费尔马大定理的一个证明，在 1995 年威尔斯又提出一个修正的证明以前，发现他原先证明中的错误便需要大量的附加性工作。大多数逻辑学家否认数学知识的一个可接受的形式化系统是不一致的，因为从一个矛盾中可以导出任何结论来。彭罗斯指出一个不一致的人类数学家将会不得不接受 1＝2 的结论。但没有理由说明为什么 CAM 的计算能力不能以更雅致的方式解决孤立的矛盾。例如，一个并行约束满足的系统可以包含不一致性，同时在大多数情况下运转良好。所以说 CAM 与人类数学家一样，可以是不一致的和不可靠的，有时会产生出错误的答案。

即使 CAM 是可靠的，要证明这一点也会是不可计算的。检查一个系统的一致性是复杂性随指数增长的一类计算的基本例子，随着问题变得越来越复杂，这类计算所耗费的时间会呈指数增长。检查少量命题之间的一致性很容易，但 n 个命题便要检查 2^n 种组合方式。如果 CAM 真是一个好的数学家模型，那么它会包含大量的数学命题，而计算其一致性将是不可能的。因而，假如 CAM 有

足够的威力来模拟人类对数学知识的理解，我们应该指望其图灵机算法不是可知为可靠的或一致的。由于 CAM 不使用可知为可靠的算法，这与人类数学家没什么明显的不同。由于彭罗斯没有证明有什么是人类数学家可以做而计算机不能做的，CRUM 可以合情合理地否认他的挑战。

扩展 CRUM

尽管如此，彭罗斯论证中存在的弱点不应当让人觉得认知科学家目前就能理解数学知识的本质。人工智能研究者已研制出了一些程序用来模拟数学思维的某些方面，比如概念的形成（Lenat 1983)、定理证明（Henschen 1990）以及类比推理。但没人敢说他能拿出一个关于数学家的演绎能力的严格的模型，更别说反映数学家的想象力了。因此，要对数学思维有进一步的了解，有必要实质地扩展目前的计算-表征的研究路线。

补充 CRUM

为了对人类的心智，包括其数学上的能力，有一个全面的了解，用新的神经生理学上的观念对 CRUM 加以补充，看来也是有可能。彭罗斯提出了一些饶有兴味的猜测，他认为在任何神经细胞中都可见的一种类似蛋白质的分子——微管，在认知和意识中可能扮演着重要的角色。他推测大脑的基本计算单元并不是神经元，像联接主义所设想的那样，而是大量的由微管所组成的**管状二聚物**。他估计每个神经元里有数千个这样的二聚物，其操作比神经元的激活要快 100 万倍。从神经元的角度看，大脑看上去是一台相对很慢的计算机，但有可能在更低的分子层次上得到更快的计算速度。在这个层次也有可能是由奇特的量子力学过程进行操作，也只有彭罗斯能猜测它们可能是什么。一般的麻醉学，其药物作用是去掉意识，可能就是从化学上中止微管的活动。也许克里克（Crick 1994）关于意识的神经学解释的推测需要由化学解

释来加以补充，要深入到神经元的内部结构并引入量子力学效应。对这一切的理解不仅需要认知科学的发展，而且还要求物理学和生物学的进步。

彭罗斯关于数学知识本质的论证未能表明心智不是计算机，而他关于大脑中的量子力学效应的猜测也还缺少证据。但我们对此应持以开放的态度，新奇的物理学和化学过程可能与理解我们大脑的运作是有关系的。

小　　结

我们可以将心智视为一个动力学系统而不是计算-表征系统。心智的动力学系统模型的出现相对较新，可以大致分为三类：使用少量的变量和方程，比较笼统地借用动力学系统的思想，或者是用动力学系统的术语来描述联接主义模型。动力学系统的方式似乎有望处理心智的时间性的、物理性的和非表征性的方面，给我们的启示是要扩展和补充CRUM，而不是抛弃它。

彭罗斯提出CRUM不能解释数学知识，特别是无法解释人类辨别无可置疑的数学真理的能力。他关于哥德尔定理驳倒了任何对心智的计算理解的论证是有毛病的，因为它预设了关于心智的计算模型是可知为可靠的。

讨 论 题

1. 将心智看成动力学系统有什么优点？

2. 动力学系统方法与CRUM是否不兼容？

3. 联接主义网络是动力学系统吗？

4. 在建构一个关于数学家的计算模型中，哪些类型的表征和加工过程是必须的？

5. 对数学家的计算模型构造一个等价的图灵机会有怎样的

困难？

6. 数学家（以及一般人）的思维符合一致性吗？

进一步的推荐读物

对动力学系统的通俗介绍，见 Gleick 1987 及 Waldrop 1992。Abraham 和 Shaw 1992 介绍了其在心理学上的应用。主张从动力学系统的路线研究认知的文献有 Smith 和 Thelen 1993，Thelen 和 Smith 1994，Van Gelder 1995 及 Van Gelder 和 Port 1995。

Hofstadter 1979 提供了对哥德尔定理的引人入胜的讨论，包括对基于哥德尔定理的反计算的论证的回击。对哥德尔定理的易懂而准确的介绍，见 DeLong 1970。对彭罗斯的早期哥德尔式论证的诸多批评，见《行为与脑科学》第 13 卷第 4 期（*Behavioral and Brain Scienes*，Vol. 13，no. 4，1990）。Grush 和 Churchland 1995 对彭罗斯关于意识的神经学思想提出了挑战。

备　　注

动力学系统一般用微分或差分方程来描述。例如，考虑如下的逻辑差分方程：

$$x_{t+1} = r * x_t * (1 - x_t)。$$

我们可以把 x_t 理解为在时刻 t 的人口，或是某种可能的比例，而把 r 解释为人口增长率。这一简单的方程可以呈现出非常奇特的行为：当 r 小于 1 时，人口量趋近于 1，当 r 在 1 到 3.57 之间时，人口量稳定或是在两个点之间来回振荡。当 r 大于 3.57 时，系统变成混沌的，产生明显的随机结果，十分敏感地依赖于 x 的起始值。可以在你的计算机或计算器上，看一看 x 的起始值是 0.4 和 0.35 所带来的结果的差异。

将 CAM 转换为图灵机可以按下面的步骤自动进行。我们可以假设 CAM 是用某种高级编程语言如 LISP 编写的。

1. 将 CAM 的 LISP 代码转换为某一特定计算机的机器语言代码。

2. 将计算机的机器代码转译成为一种称为随机存贮机器的抽象装置的

代码。

3. 将随机存贮机器的代码转译为图灵机代码。

原则上，每一步都是可计算的，但最终的图灵机代码的惊人的长度和复杂性看来超出了人类的理解力。

第十二章　认知科学的未来

整　　合

尽管对心智的计算-表征观的各方面的挑战表明它需要加以扩展和补充，我们不应当忘记CRUM所取得的各种不凡的成就。正如在第八章中评述过的那样，对心智的任何其他的研究途径必须能超越CRUM在解释人类的问题求解、学习和语言时所展现出来的能力。在过去的30多年里，认知科学家在理解人类思维的诸多方面取得了可喜的进展。但是，在有些方面仍显得特别地令人难以捉摸，比如我们在第九章中讨论过的对情绪和意识的思考。我们不能指望对这些难题的解决会集中于单个学科内的进展，而应当依靠心理学、计算、神经科学、哲学、语言学和人类学等学科研究的整合，这是认知科学中最富成效的研究方针。揭示心智是如何工作的可以说是人类试图拼接的一个最大的字画谜，而各个分块需要来自多个领域的努力。

从目前的研究状况看，我认为我们可以继续推进三种类型的整合。首先，在最一般的、概念的层面上，我们实现新的跨学科的整合，心理学、哲学、人工智能、人类学、语言学和神经科学的研究人员进一步认识到相互间进行对话和交流的必要性。有志于探究心智的学生和研究人员如果局限于单个学科的理论和方法所失去的不仅仅是对心智的更广泛的理解，而且还可能丧失从跨学科的交流中激励出在他们本学科中取得创造性成果的可能性。

其次，通过利用不同学科的方法所收集到的不同类型的数据，认知科学还应当进一步推进实验上的整合。例如，对语言的研究

就需要将定性的语言学数据与实验心理学和神经学的数据合起来考虑。对表象的研究为如何将行为实验与神经学实验结合起来以支持关于心理过程的理论提供了一个很好的案例（Kosslyn 1994），对思维的其他方面的研究也同样从行为实验和神经学实验的结合中获益匪浅（Posner 和 Raichle 1994）。要解决意识这一异常困难的问题不仅要求重视行为学和神经学上的数据，还要兼顾通过我们每人的意识经验得来的实验数据。

认知科学应当进一步加强的第三种整合是由计算思想与模拟所带来的理论上的整合。我们已经看到了思维与计算之间的类比在两个方面为我们理解心智作出了贡献。第一，它对理解心理结构与过程提供了丰富的思想武器，由此能作出对心智的复杂的说明，既避免了行为主义在解释上的贫困，又躲开了二元论对心智的神秘化。较之以往的任何其他的理论途径，将心智视为一台计算机并由此推测其如何进行编程的，可以使研究者对心理的运作机制作出更为精确和详尽的说明。第二，由于计算假说能够足以精确地进行编程，通过运行模拟来进行测试，其性能便可与人类的思维行为进行对比。心智的计算观的结果之一便是人们认识到思维是何等复杂和多样化的过程；通过模拟可以让研究者们看到他们的理论思路的成就及其局限。

越来越多地，计算理论正跨越我在前面分别介绍的六种基本路线的界线而实现整合。例如，勒纳德的 CYC 工程将基于逻辑和基于概念的表征进行结合（Lenat 和 Guha 1990）。基于规则的和联接主义的方法也在开发对基于规则的系统进行联接主义式的实现的工作中得以整合（Touretzky 和 Hinton 1988）。最近许多研究类比的工作都采用了混合型的视角，结合了传统人工智能的观念如逻辑和规则以及联接主义关于并行约束满足的思想（Holyoak 和 Barnden 1994）。桑（Sun 1994）研制了一个分布式表征与局部式表征进行结合的计算模型，以便产生基于规则和基于相似的推理。表象既可以用联接主义式的、并行约束满足的词汇进行讨论

(Kosslyn 1994)，也可以采用面向逻辑的术语（Glasgow 1993）。

这样的交叉互助和融合，也许还能通过引入神经学和动力学系统的思想而得到进一步的丰富，这将成为认知科学发展的趋势。我们还可以期待认知现象所发生的物理学的、生物学的和社会性的背景会受到越来越多的重视。CRUM 正在被扩展为 CRUMBS，即生物学-社会性的，对心智的计算-表征理解。也许甚至像意识的本质这样困难的问题也会在一种整合性、多学科研究的攻击下得以解决。

你在认知科学中的未来

对认知科学的挑战以及认知科学的前景使我们清楚地看到这里有很多生机勃勃、充满希望的研究前沿。但对于有志于探索心智的大学生们来说，面临着一份困难的抉择，因为可以从许多不同学科的视角和方法来开展对心智的研究。作为大学生还可以先作一下浏览，从认知心理学、人工智能、心智哲学、认知人类学、语言学和认知神经科学中选择相关的课程听一听。而研究生则要更侧重于专门的方向。下面是为面临选择如何继续学习和研究心智本质的同学提供的一些坦率的建议：

1. 挑选出心智最让你感兴趣的一个方面。你觉得思维的哪些方面对你最有吸引力？

2. 挑选一种方法论。哪一种研究最适合你的兴趣和天分？你觉得进行实验设计还是计算机编程更引人入胜？你是喜欢收集语言学案例，还是喜欢像哲学家那样思考思维的规范性方面的问题？方法论上的选择对于你进一步寻求哪方面的训练较之别的东西更具有决定性。

3. 注意留心其他领域内的进展。在选定了认知科学中的某一分支学科后，你所面临的危险是为了胜任某一已设定的领域内的

研究而陷入繁重的事务中，从而忽视了与其他领域之间的联系。目前只有一部分大学设有认知科学的研究生培养规划（参见本章附录所提供的指向诸多认知科学研究生培养规划的全球信息网（World Wide Web）的入口）。在传统的狭窄专业就读的学生必须付出更多的努力以保持住多学科性的兴趣。

4. 整合与协作。避免独断：不要认为在你本系占主导地位的理论和方法论的途径是研究心智的唯一方式。除了寻求理论上的整合以外，对于方法论上的综合也应持开放的态度——例如，既进行实验研究也要开展计算机模拟。由于即便是掌握一种基本方法也是不简单的任务，可能占用你整个研究生的学习阶段，因此可以寻找一些具有共同兴趣却有不同技能的合作者。在认知科学中不少优异的工作都是由综合了不同的学识见解和方法论的研究者合作完成的。

不打算深入认知科学研究的学生也有充足的动因对这一领域保持关注，因为认知科学具有进一步实践应用的前景。法律、医学、工程、商业、艺术和教育都是与改善对心智的理解有密切联系的领域。

在第一章的结尾处，我提出了指导本书写作的六个基本假定，而现在我希望这一切都已得以兑现了。认知科学是一项引人入胜、激动人心的事业，涉及多种多样的、跨学科的研究路线，其核心是对心智的计算-表征理解，而这一核心还需要进一步加以扩展和补充。在认知科学中有许许多多令人激动的项目等待着未来的探索者。

小　　结

认知科学无论是在理论上还是在应用上都已经取得了可喜的成就，但对进一步的理论和实验发展仍留有巨大的用武之地。要

在对心智的理解上取得进展不能局限于狭窄的领域，而要求跨学科的、实验上的和理论上的综合。对面向认知科学未来的同学而言，有大量的未解之谜和不同的研究途径可供选择。

讨论题

1. 对于理解心智的本质，认知科学做出了多少贡献？还有多少问题尚待解决？哪一方面给你的印象更深？

2. 在认知科学中进行进一步的实验和理论上的整合，还存在什么障碍？

3. 你对认知科学中什么研究方法最有兴趣？什么方法对今后探索心智本质最有效？

进一步的推荐读物

参见第一章的末尾关于认知科学的一般性读物。也可参阅 Holyoak 和 Spellman 1993。Becatel 1986 讨论了不同科学学科进行综合所面临的问题。

附录：认知科学的资源

注：这个目录不是完整详尽的，只求对一些最有用的资源提供导引。

工具书

Dictionary of Psychology	心理学词典
Dictionary of Philosophy	哲学词典
Encyclopedia of Artificial Intelligence	人工智能百科全书
Encyclopedia of Philosophy	哲学百科全书
Glossary of Cognitive Science	认知科学词汇表
International Dictionary of Psychology	国际心理学词典

期　刊

跨学科性的

Behavioral and Brain Sciences	行为科学与脑科学
Cognition	认知
Cognitive Science	认知科学
Journal of the Learning Sciences	学习科学杂志
Linguistics and Philosophy	语言学与哲学
Mind and Language	心智与语言

哲　学

Journal of Philosophy	哲学杂志
Mind	心智
Minds and Machines	心智与机器
Philosophical Psychology	哲学心理学

心理学

Cognition and Emotions	认知与情绪
Cognition and Instruction	认知与教学
Cognitive Psychology	认知心理学
Journal of Experimental Psychology: Learning, Memory, and Cognition	实验心理学杂志:学习、记忆与认知
Psychological Review	心理学评论

人工智能

Artificial Intelligence	人工智能
Computational Intelligence	计算智能
Connection Science	联接科学
Journal of Theoretical and Applied Artificial Intelligence	理论与应用人工智能杂志
Machine Learning	机器学习

神经科学

Cognitive Neuroscience	认知神经科学
Neural Networks	神经网络

语言学

Foundations of Language	语言基础
Language	语言
Linguistic Inquiry	语言学研究

人类学和社会学

Current Anthropology	当代人类学
Social Studies of Science	科学的社会研究

学术组织

American Association for Artificial Intelligence 全美人工智能学会

Cognitive Neuroscience Society	认知神经科学学会
Cognitive Science Society	认知科学学会
Society for Machines and Mentality	机器与心智研究会
Society for Philosophy and Psychology	哲学与心理学研究会

会议论文集

Advances in Neural Information Processing systems 神经信息加工系统进展，由 Morgan Kaufmann 出版社出版

Proceedings of the Cognitive Science Society Conference 认知科学学会学术大会论文集，由 Erlbaum 出版社出版

Proceedings of the International Joint Conference in Artificial Intelligence 国际人工智能大会论文集，由 Morgan Kaufmann 出版社出版

Proceedings of the National Conference on Artificial Intelligence 全国人工智能大会论文集，由 AAAI 出版社出版并由 MIT 出版社发行

出版社

经常出版与认知科学有关书籍的出版社有：Basil Blackwell Ltd. Cambridge University Press，Harvard Universi:ty Press，Kluwer Academic Publishers，Lawrence Erlbaum Associates，MIT Press，Morgan Kaufmann Publishers，Universty of Chicago Press。

Internet

在 WWW(Word wide web，全球信息网)上，斯坦福(Stanford)大学提供的一个网页上含有许多关于各种认知科学研究生培养规划的信息，其 URL(统一资源地址标识)是：http://www—psych.stanford.edu/cogsci/. 另外可参见 http://www.cog.brown.edu/pointers/cognitive.html.

人工智能的资源可以通过 http://ai.iit.nrc.ca/misc.html. 找到。

我在滑铁卢(Waterloo)大学的实验室的 URL 是 http://cogsci.uwaterloo.ca/.

参 考 文 献

Abraham, R. H.", and C. D. Shaw, 1992. *Dynamics: The geometry of behavior*. 2nd ed. Redwood City, Calif.: Addison-Wesley.

Adams, M. J. 1990. *Beginning to read*. Cambridge, Mass.: MIT Press.

Aitchison, J. 1987. *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.

Akmajian, A., R. A. Demers, A. K. Farmer, and R. M. Harnish. 1995. *Linguistics: An introduction to language and communication*. 4th ed. Cambridge, Mass.: MIT Press.

Allen, B. P. 1994. Case-based reasoning: Business applications. *Communications of the ACM*, 37(March), 40-42.

Allen, R. H., ed. 1992. *Expert systems for civil engineers: Knowledge representation*. New York: American Society of Civil Engineers.

Anderson, J. A., and E. Rosenfeld, ed. 1988. *Neurocomputing*. Cambridge, Mass.: MIT Press.

Anderson, J. R. 1983. *The architecture of cognition*. Cambridge, Mass.: Harvard University Press.

Anderson, J. R. 1990. *Cognitive science and its implications*. 3rd ed. New York: Freeman.

Anderson, J. R. 1993. *Rules of the mind*. Hillsdale, N. J.: Erlbaum.

Baars, B. J 1988. *A cognitive theory of consciousness*. Cambridge: Cambridge University Press.

Barnes, A., and P. Thagard. In press. Empathy and analogy. *Dialogue*.

Barrett, R. A. 1991. *Culture and conduct: An excursion in anthropology*. 2nd ed. Belmont, Calif.: Wadsworth.

Barsalou, L. W. 1983. Ad hoc categories. *Memory and Cognition* 11, 211—227.

Bartlett, F. C. 1932. *Remembering*. Cambridge: Cambrideg University Prss.

Barton, S. 1994. Chaos, self-organization, and psychology. *American Psychologist* 49, 5-14.

Bechtel, W., ed. 1986. *Integrating scientific disciplines*. Dordrecht: Kluwer.

Bechtel, W. 1988. *Philosophy of mind: An introduction for cognitive science*. Hillsdale, N. J.: Erlbaum.

Bechtel, W. 1993. Currents in connectionism. *Minds and Machines* 3, 125-153.

Bechtel, W., and A. Abrahamsen. 1991. *Connectionism and the mind*. Oxford: Blackwell.

Benacerraf, P. 1967. God, the devil, and Gödel. *The Monist* 51, 9-32.

Block, N. 1978. Troubles with functionalism. In C. W. Savage, ed., *Perception and cognition*. Minneapolis, Minn.: University of Minnesota Press.

Boden, M. 1988. *Computer models of mind*. Cambridge: Cambridge University Press.

Bond, A., and L. Gasser, eds. 1988. *Readings in distributed artificial intelligence*. San Mateo, Calif.: Morgan Kaufmann.

Braine, M. D. S. 1978. On the relation between the natural logic of reasoning and standard logic. *psychological Review* 85, 1-21.

Brewer, W., and J. Treyens. 1981. Role of schemata in memory for places. *Cognitive Psychology* 13, 207-230.

Brooks, R. A. 1991. Intelligence without representation. *Artificial Intelligence* 47, 139-159.

Bruer, J. T. 1993. *Schools for thought: A science of learning in the classroom*. Cambridge, Mass.: MIT Press.

Bruner, J. S. 1990. *Acts of meaning*. Cambridge, Mass.: Harvard U-

niversity Press.

Bruner, J. S., J. J. Goodnow, and G. A. Austin. 1956. *A study of thinking*. New York: Wiley.

Buchanan, B., and E. Shortliffe, eds. 1984. *Rule-based expert systems*. Reading, Mass.: Addison-Wesley.

Busemeyer, J. R., and J. T. Townsend. 1993. Decision field theory: A dynamiccognitive approach to decision making in an uncertain environment. *Psychological Review* 100, 432 459.

Card, S. K., T. P. Moran, and A. Newell. 1983. *The psychology of computer-human interaction*. Hillsdale, N. J.: Erlbaum.

Cheng, P. W., and K. J. Holyoak. 1985. Pragmatic reasoning schemas. *Cognitive Psychology* 17, 391-416.

Chi, M. 1992. Conceptual change within and across ontological categories: Examples from learning and discovery in science. In R. Giere, ed., *Cognitive models of science*, 129—186. Minneapolis, Minn.: University of Minnesota Press.

Chomsky, N. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, N. 1959. A review of B. F. Skinner's *Verbal behavior*. *Language* 35, 26-58.

Chomsky, N. 1972. *Language and mind*. 2nd ed. New York: Harcourt Brace Jovanovich.

Chomsky, N. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, N. 1988. *Language and problems of knowledge*. Cambridge, Mass.: MIT Press.

Churchland, P. M. 1989. *A neurocomputational perspective*. Cambridge, Mass.: MIT Press.

Churchland, P. M. 1995. *The engine of reason, the seat of the soul*. Cambridge, Mass.: MIT Press.

Churchland, P. S. 1986. *Neurophilosophy*. Cambridge, Mass.: MIT Press.

Churchland, P. S., and T. Sejnowski. 1992. *The computationl brain.* Cambridge, Mass.: MIT Press.

Cooper, L. A., and R. N. Shepard. 1973. Chronometric studies of the rotation of mental images. In W. G. Chase, ed., *Visual information processing*, 75--176. New York: Academic Press.

Copi, I. 1979. *Symbolic logic.* 5th ed. New York: Macmillan.

Crick, F. 1994. *The astonishing hypothesis: The scientific search for the soul.* London: Simon and Schustet.

Crowley, K., and R. S. Siegler. 1993. Flexible strategy use in children's tic-tac toe. *Cognitive Science* 17, 531—561.

Damasio, A. R. 1994. *Descartes' error.* New YorK: Putnam.

D' Andrade, R. G. 1995. *The development of cognitive anthropology.* Cambridge: Cambridge University Press.

Dean, T. L., and M. P. Wellman. 1991. *Planning and control.* San Mateo, Calif.: Morgan Kaufmann.

DeLong, H. 1970. *A profile of mathematical logic.* Reading, Mass.: AddisonWesley.

Dennett, D. 1991. *Consciousness explained.* Boston: Little, Borwn.

Dietrich, E., ed. 1994. *Thinking computers and virtual persons: Essays on the intentionality of machines.* San Diego, Calif.: Academic Press.

Dreyfus, H. L. 1991. *Being-in-the-world.* Cambridge, Mass.: MIT Press.

Dreyfus, H. L. 1992. *What computers still can't do.* 3rd ed. Cambridge, Mass.: MIT Press.

Durfee, E. 1992. What your computer really needs to know, you learned in kindergarten. In *AAAI-92: Proceedings of the Tenth National Conference on Artificial Intelligence*, 858—864. Menlo Park, Calif.: AAAI Press; distributed by MIT Press.

Durfee, E., V. Lesser, D. Corkhill. 1989. Cooperative distributed problem solving. In A. Barr, P. Cohen, and E. Feigenbaum, eds., *The handbook of artificial intelligence.* Vol. IV., 83—147. Reading, Mass.:

Addison-Wesley.

Dym. C. L., and R. E. Levitt. 1991. *Knowledge-based systems in engineering*. New York: McGraw-Hill.

Edelman, G. M. 1992. *Bright air, brilliant fire: On the matter of the mind*. New York: Basic Books.

Evans, T. 1968. A program for the solution of a class of geometric analogy intelligence test questions. In M. Minsky, ed., *Semantic information processing*, 271- 353. Cambridge, Mass.: MIT Press.

Feigenbaum, E., P. McCorduck, and H. Nii. 1988. *The rise of the expert company*. New York: Vintage.

Feldman, J. A. 1981. A connectionist model of visual memory. In G. E. Hinton and J. A. Anderson, eds., *Parallel models of associative memory*, 49—81. Hillsdale, N. J.: Erlbaum.

Fikes, R., and N. Nilsson. 1971, STRIPS: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2, 189—208.

Finke, R. 1989. *Principles of mental imagery*. Cambridge, Mass.: MIT Press.

Finke, R., S. Pinker, and M. Farah. 1989. Reinterpreting visual patterns in mental imagery. *Cognitive Science* 13, 51—78.

Finke, R., T. B. Ward, and S. M. Smith. 1992. *Creative cognition: Theory, research, and applications*. Cambridge, Mass.: MIT Press.

Flanagan, O. 1992. *Consciousness reconsidered*. Cambridge, Mass.: MIT Press.

Flanagan, O. 1995. Deconstructing dreams. The spandrels of sleep. *Journal of Philosophy* 92, 5—27.

Fodor, J. 1975. *The language of thought*. New York: Crowell.

Fodor, J. 1987. *Psychosemantics*. Cambridge, Mass.: MIT Press.

Forbus, K., D. Gentner, and K. Law. 1995. MAC/FAC: A model of similaritybased retrieval. *Cognitive Science* 19, 144—205.

Forbus, K., P. Nielsen, and B. Faltings. 1991. Qualitative spatial

reasoning: The CLOCK project. *Artificial Intelligence* 51, 417—472.

Foss, J. E. 1995. Materialism, reduction, replacement, and the place of consciousness in science. *Journal of Philosophy* 92, 401—429.

Frege, G. 1960. *Translations from the philosophical writings of Gottlob Frege*. Trans. by P. Geach and M. Black. Oxford: Basil Blackwell.

Frijda, N. H. 1986. *The emotions*. Cambridge: Cambridge University Press.

Funt, B. 1980. Problem solving with diagrammatic representations. *Artificial Intelligence* 13, 201—230.

Galegher, J., R. E. Kraut, and C. Egido, eds. 1990. *Intellectual teamwork: Social and technological foundations of cooperative work*. Hillsdale, N. J.: Erlbaum.

Gardner, H. 1985. *The mind's new science*. New York: Basic Books.

Gasser, L. 1991, Social conceptions of Knowledge and action: DAI and open systems semantics. *Artificial Intelligence* 47, 107—138.

Genesereth, M. R., and N. J. Nilsson. 1987. *Logical foundations of artificial intelligence*. Los Altos, Calif.: Morgan Kaufmann.

Gentner, D. 1983. Structure-mapping: A theoretical framework for analogy. *Cognitive Science* 7, 155—170.

Gentner, D. 1989. The mechanisms of analogical learning. In S. Vosniadou and A. Ortony, eds., *Similarity and analogical reasoning*, 199—241. Cambridge: Cambridge University Press.

Gibson, J. J. 1979. *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gick, M. L., and K. J. Holyoak. 1980. Analogical problem solving. *Cognitive Psychology* 12, 306—355.

Gich, M. L., and K. J. Holyoak. 1983. Schema induction and analogical transfer. *Cognitive Psychology* 15, 1—38.

Giere, R. 1988. *Explaining science: A cognitive approach*. Chicago: University of Chicago Press.

Gigerenzer, G., U. Hoffrage, and H. Kleinbölting. 1991. Probabilis-

tic mental models: A Brunswikian theory of confidence. *Psychological Review* 98, 506—528.

Glasgow, J. I. 1993. The imagery debate revisited: A computational perspective. *Computational Intelligence* 9, 309—333.

Glasgow, J. I., S. Fortier, and F. Allen. (1993). Molecular scene analysis: Crystal structure determination through imagery. In L. Hunter, ed., *Artificial intelligence and molecular biology*, 433—458. Cambridge, Mass.: MIT Press.

Glasgow, J. I., and D. Papadias. 1992. Computational imagery. *Cognitive Science* 16, 355—394.

Gleick, J. 1987. *Chaos: Making a new science.* New York: Viking.

Glucksberg, S., and B. Keysar. 1990. Understanding metaphorical comparisons: Beyond similarity. *Psychological Review* 97, 3—18.

Goldman, A. I. 1992. *Liaisons: Philosophy meets the cognitive and social sciences.* Cambridge, Mass.: MIT Press.

Goleman, D. 1995. *Emotional intelligence.* New York: Bantam.

Goss, S., C. Hall, E. Buckolz, and G. Fishburne. 1986. Imagery ability and the acquisition and retention of movements. *Memory and Cognition* 14, 469—477.

Graham, G. 1993. *Philosophy of mind: An introduction.* Oxford: Blackwell.

Gray, J. A. In press. The contents of consciousness: A neuropsychological conjecture. *Behavioral and Brain Sciences.*

Grush, R., and P. S. Churchland. 1995. Gaps in Penrose's toiling. *Journal of Consciousness Studies* 2, 10—29.

Hall, R. 1989. Computational approaches to analogical reasoning: A comparative analysis. *Artificial Intelligence* 39, 39—120.

Hebb, D. O. 1949. *The organization of behavior.* New York: Wiley.

Heidegger, M. 1962. *Being and time.* Trans. by J. Macquarrie and E. Robinson. New York: Harper & Row.

Hempel, C. G. 1965. *Aspects of scientific explanation.* New York:

Free Press.

Henschen, L. 1990. Theorem proving. In S. C. Shapiro, ed., *Encyclopedia of artificial intelligence*, 1114–1123. New York: Wiley.

Hesse, M. 1966. *Models and analogies in science.* Notre Dame, Ind.: Notre Dame University Press.

Hewitt, C. 1991. Open information systems semantics for distributed artificial intelligence. *Artificial Intelligence* 47, 79–106.

Hinkle, D., and C. N. Toomey. 1994. Clavier: Applying case-based reasoning in composite part fabrication. In *Proceedings of the Sixth Innovative Applications of Artificial Intelligence Conference*, 54–62. Menlo Park, Calif.: AAAI Press.

Hinton, G. E. 1990. Connectionist learning procedures. In J. Carbonell, ed., *Machine learning: Paradigms and methods*, 185–234. Cambridge, Mass.: MIT Press.

Hinton, G. E. and J. A. Anderson, eds. 1981. *Parallel models of associative memory.* Hillsdale, N. J.: Erlbaum.

Hirschfeld, L. A., and S. A. Gelman, eds. 1994. *Mapping the mind: Domain specificity in cognition and culture.* Cambridge: Cambridge University Press.

Hofstadter, D. 1979. *Gödel, Escher, Bach: An eternal golden braid.* New York: Basic Books.

Hofstadter, D. 1995. *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought.* New York: Basic Books.

Holland, J. H., K. J. Holyoak, R. E. Nisbett, and P. R. Thagard. 1986. *Induction: Processes of inference, learning, and discovery.* Cambridge, Mass.: MIT Press.

Holtzman, S. 1989. *Intelligent decision systems.* Reading, Mass.: Addison-Wesley.

Holyoak, K. J., and J. A. Barnden, eds. 1994. *Advances in connectionist and neural computational theory.* Vol. 2, *Analogical connections.* Norwood, N. J.: Ablex.

Holyoak, K. J., and B. A. Spellman. 1993. Thinking. *Annual Review of Psychology* 44, 265--315.

Holyoak, K. J., and P. Thagard. 1995. *Mental leaps: Analogy in creative thought.* Cambridge, Mass.: MIT Press.

Howson, C., and P. Urbach. 1989. *Scientific reasoning: The Bayesian tradition.* Lasalle, Ill.: Open Court.

Hutchins, E. 1995. *Cognition in the wild.* Cambridge, Mass.: MIT Press.

Jackendoff, R. 1987. *Consciousness and the computational mind.* Cambridge, Mass.: MIT Press.

Johnson, M. 1987. *The body in the mind.* Chicago: University of Chicago Press.

Johnson-Larid, P. N. 1983. *Mental models.* Cambridge, Mass.: Harvard University Press.

Johnson-Laird, P. N. 1988. *The computer and the mind.* Cambridge, Mass.: Harvard University Press.

Johnson-Laird, P. N., and R. M. Byrne. 1991. *Deduction.* Hillsdale, N. J.: Erlbaum.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kant, I. 1965. *Critique of pure reason.* Trans. by N. Kemp Smith. 2nd ed. London: Macmillan.

Keil, F. 1989. *Concepts, kines, and cognitive development.* Cambridge. Mass.: MIT Press.

Keysar, B. 1990. On the functional equivalence of literal and metaphorical interpretations in discourse. *Journal of Memory and Language* 28, 375—385.

Kintsch, W. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review* 95, 163-182.

Kintsch, W., D. Welsch, F. Schmalhofer, and S. Zimny. 1990. Sentence memory: A theoretical analysis. *Journal of Memory and Language*

29, 133—159.

Kitayama, S., and H. R. Markus, eds. 1994. *Emotion and culture: Empirical studies of mutual influence*. Hyattsville, Md.: American Psychological Association.

Kitcher, P. 1981. Explanatory unification. *Philosophy of Science* 48, 507-531.

Kitcher, P. 1993. *The advancement of science*. Oxford: Oxford University Press.

Kolodner, J. 1993. *Case-based reasoning*. San Mateo, Calif.: Morgan Kaufmann.

Konolige, K. 1992. Abduction versus closure in causal theories. *Artificial Intelligence* 53, 255—272.

Kosslyn, S. M. 1980. *Image and mind*. Cambridge, Mass.: Harvard University Press.

Kosslyn, S. M. 1994. *Image and brain: The resolution of the imagery debate*. Cambridge, Mass.: MIT Press.

Kosslyn, S. M., and O. Koenig. 1992. *Wet mind: The new cognitive neuroscience*. New York: Free Press.

Kosslyn, S. M., and S. P. Shwartz. 1977. A simulation of visual imagery. *Cognitive Science* 1, 265—295.

Kramer, P. D. 1993. *Listening to Prozac*. New York: Viking.

Kunda, Z., D. Miller, and T. Claire. 1990. Combining social concepts: The role of causal reasoning. *Cognitive Science* 14, 551—577.

Kunda, Z., and P. Thagard. 1996. Forming impressions using stereotypes, traits, and behaviors: A parallel constraint satisfaction. *theory Psychological Review* 103, 284—308.

Lakoff, G. 1987. *Women, fire, and dangerous things*. Chicago: University of Chicago Press.

Lakoff, G. 1994. What is metaphor? In J. A. Barnden and K. J. Holyoak, eds., *Advances in connectionist and neural computation theory*. Vol. 3, *Analogy, metaphor, and reminding*, 203-257. Norwood, N. J.:

Ablex.

Lakoff, G., and M. Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Langacker, R. W. 1987. *Foundations of cognitive grammar*. Stanford, Calif.: Stanford University Press.

Langley, P., and H. A. Simon. 1995. Applications of machine learning and rule induction. *Communications of the ACM* 38(November), 55—64.

Larkin, J. H., and H. A. Simon. 1987. Why a diagram is (sometimes)worth ten thousand words. *Cognitive Science* 11, 65—100.

Latour, B., and S. Woolgar. 1986. *Laboratory life: The construction of scientific facts*. Princeton, N. J.: Princeton University Press.

Lave, J., and E. Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Leake, D. B. 1992. *Evaluating explanations: A content theory*. Hillsdale. N. J.: Erlbaum.

LeDoux, J. E. 1993. Emotional networks in the brain. In M. Lewis and J. M. Haviland, eds., *Handbook of emotions*, 109—118. New York: Guilford Press.

Lenat, D. 1983. The role of heuristics in learning by discovery: Three cast studies. In R. Michalski, J. Carbonell, and T. Mitchell, eds., *Machine learning: An artificial intelligence approach*, 243—306. Palo Alto, Calif.: Tioga.

Lenat, D., and R. Guha. 1990. *Building large knowledge-based systems*. Reading, Mass.: Addison-Wesley.

Levine, D. S. 1991. *Introduction to neural and cognitive modeling*. Hillsdale, N. J.: Erlbaum.

Lewis, M., and J. M. Haviland, eds. 1993. *Handbook of emotions*. New York: Guilford Press.

Ling, C., and M. Marinov. 1993. Answering the connectionist challenge: A symbolic model of learning past tenses of English verbs. *Cognition*

49, 235—290.

Lucas, J. R. 1961. Minds, machines, and Gödel. *Philosophy* 36, 120—124.

Makhworth, A. 1993. On seeing robots. In A. Basu and X. Li, eds., *Computer vision: Systems, theory, and applications*, 1—13. Singapore: World Scientific.

MacWhinney, B., and J. Leinbach. 1991. Implementations are not conceptualizations: Revising the verb model. *Cognition* 40, 121—157.

Maida, A. S. 1990. Frame theory. In S. C. Shapiro, ed., *Encyclopedia of artificial intelligence*, 302—312. New York: Wiley.

Mannes, S. M., and W. Kintsch. 1991. Routine computing tasks: Planning as understanding. *Cognitive Science* 15, 305—342.

Marr, D. 1982. *Vision*. San Francisco: Freeman.

Marr, D., and T. Poggio. 1976. Cooperative computation of stereo disparity. *Science* 194, 283—287.

McCawley, J. D. 1993. *Everything that linguists have always wanted to know about logic-but were ashamed to ask.* 2nd ed. Chicago: University of Chicago Press.

McClelland, J. L., and J. L. Elman. 1986. The TRACE model of speech perception. *Cognitive Psychology* 18, 1—86.

McClelland, J. L., and D. E. Rumelhart. 1981. An interactive activation model of context effects in letter perception. Part 1, An account of basic findings. *Psychological Review* 88, 375—407.

McClelland, J. L., and D. E. Rumelhart. 1989. *Explorations in parallel distributed processing.* Cambridge, Mass.: MIT Press.

Medin, D. L., and B. H. Ross. 1992. *Cognitive psychology.* Fort Worth, Tex.: Harcourt Brace Jovanovich.

Michalski, R., J. Carbonell, and T. Mitchell, eds. 1986. *Machine learning: An artificial intelligence approach.* Vol. 2. Los Altos, Calif.: Morgan Kaufmann.

Mill, J. S. 1974. *A system of logic ratiocinative and inductive.* Toron-

to: University of Toronto Press.

Miller, G. A. 1956. The magical number seven, plus or minus two. Some limits on our capacity for processing information. *Psychological Review* 63, 81—97.

Miller, G. A, R. Bechwith, C. Fellbaum, G. Gross, and K. Miller. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography* 3, 235—244.

Miller, G. A. 1991. *The science of words*. New York: Scientific American Library.

Minsky, M. 1975. A framework for representing knowledge. In P. H. Winston, ed., *The psychology of computer vision*, 211—277. New York: McGraw-Hill.

Mitchell, M. 1993. *Analogy-making as perception*. Cambridge, Mass.: MIT Press.

Montague, R. 1974. *Formal philosophy: Selected papers of Richard Montague*. New Haven, Conn.: Yale University Press.

Murphy, G., and D. L. Medin. 1985. The role of theories in conceptual coherence. *Psychological Review* 92, 289—316.

Nagel, E., and J. R. Newman. 1958. *Gödel's proof*. London: Routledge and Kegan Paul.

Neapolitain, R. 1990. *Probabilistic reasoning in expert systems*. New York: Wiley.

Nelson, G., P. Thagard, S. Hardy. 1994. Integrating analogies with rules and explanations. In K. J. Holyoak and J. A. Barnden, eds., *Advances in connectionist and neural computational theory*. Vol. 2, *Analogical connections*, 181 205, Norwood, N. J.: Ablex.

Nersessian, N. 1989. Conceptual change in science and in science education. *Synthese* 80, 163—183.

Newell, A. 1990. *Unified theories of cognition*. Cambridge, Mass.: Harvard University Press.

Newell, A., J. C. Shaw, and H. A. Simon, 1958. Elements of a the-

ory of human problem solving. *Psychological Review* 65, 151—166.

Newell, A., and H. A. Simon. 1972. *Human problem solving*. Englewood Cliffs, N. J.: Prentice-Hall.

Nisbett, R. E., ed. 1993. *Rules for reasoning*. Hillsdale, N. J.: Erlbaum.

Norman, D. A. 1989. *The design of everyday things*. New York: Doubleday.

Oatley, K. 1992. *Best laid schemes: The psychology of emotions*. Cambridge: Cambridge University Press.

Oatley, K., and E. Duncan. 1994. The experience of emotions in everyday life. *Cognition and Emotion* 8, 369—381.

Oatley, K., and L. Larocque. 1995. Everyday concepts of emotions: Following every-other-day errors in joint plans. Ms., In J. Russell, J. M. Fernandez-Dols, A. S. R. Manstead, and J. Wellenkamp, eds., *Everyday conceptions of emotions: An introduction to the psychology, anthropology, and linguistics of emotion*, 145—165. Dordrecht: Kluwer.

O'Brien, D. P., M. D. S. Braine, and Y. Yang. 1994. Propositional reasoning by mental models? Simple to refute in principle and in practice. *Psychological Review* 101, 711—724.

Ortony, A., G. L. Clore, and A. Collins. 1988. *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Osherson, D. N. 1995. *An invitation to cognitive science*. 3 vols. 2nd ed. Cambridge, Mass.: MIT Press.

Paivio, A. 1971. *Imagery and verbal processes*. New York: Holt Rinehart and Winston.

Pearl. J. 1988. *Probabilistic reasoning in intelligent systems*. San Mateo, Calif.: Morgan Kaufmann.

Peirce, C. S. 1992. *Reasoning and the logic of things*. Cambridge, Mass.: Harvard University Press.

Penrose, R. 1989. *The emperor's new mind: Concerning computers, minds, and the laws of physics*. Oxford: Oxford University Press.

Penrose, R. 1994. *Shadows of the mind: A search for the missing science of consciousness.* Oxford: Oxford University Press.

Piaget, J., and B. Inhelder. 1969. *The psychology of the child* Trans. by H. Weaver. New York: Basic Books.

Pinker, S. 1994. *The language instinct: How the mind creates language.* New York: Morrow.

Pinker, S., and A. Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 28, 73—193.

Pollack, J. B. 1990. Recursive distributed representations. *Artificial Intelligence* 46, 77—105.

Pollack, J. B. 1991. The induction of dynamical recognizers. *Machine Learning* 7, 227—252.

Pollock, J. B. 1989. *How to build a person: A prolegomenon.* Cambridge, Mass.: MIT Press.

Polya, G. 1957. *How to solve it.* Princeton, N. J.: Princeton University Press.

Port, R., and T. van Gelder, eds. 1995. *Mind as motion: Explorations in the dynamics of cognition.* Cambridge, Mass.: MIT Press.

Posner, M. I., ed. 1989. *Foundations of cognitive science.* Cambridge, Mass.: MIT Press.

Posner, M. I., and S. W. Keele. 1970. Retention of abstract ideas. *Journal of Experimental Psychology* 83, 304—308.

posner, M. I., and M. E. Raichle. 1994. *Images of mind.* New York: Freeman.

Prince, A., and P. Smolensky. To appear. *Optimality Theory: Constraint interaction in generative grammar.* Cambridge, Mass.: MIT Press.

Prior, A. N. 1967. Logic, history of. In P. Edwards, ed., *Encyclopedia of philosophy*, 513 571. New York: Macmillan.

Putnam, H. 1975. *Mind, language, and reality.* Cambridge: Cambridge University Press.

Pylyshyn, Z. 1984. *Computation and cognition: Toward a foundation for cognitive science.* Cambridge, Mass. : MIT Press.

Quinlan, J. R. 1983. Learning efficient classification procedures and their application to chess end games. In R. Michalski, J. Carbonell, and T. Mitchell, eds., *Machine learning: An artificial intelligence approach*, 463—482. Palo Alto, Calif. : Tioga.

Read, S., and A. Marcus-Newhall. 1993. The role of explanatory coherence in the construction of social explanations. *Journal of Personality and Social Psychology* 65, 429—447.

Resnick, L., J. Levine, and S. Behrend, eds. 1991, *Socially shared cognitions.* Hillsdale, N. J. : Erlbaum.

Rich, E., and K. Knight. 1991. *Artificial intelligence.* 2nd ed. New York: McGraw-Hill.

Richards, D. 1990. Is strategic decision making chaotic? *Behavioral Science.* 35, 219 232.

Riesbeck, C. K., and R. C. Schank. 1989. *Inside case-based reasoning.* Hillsdale, N. J. : Erlbaum.

Rips, L. J. 1983. Cognitive processes in propositional reasoning. *Psychological Review* 90, 38—71.

Rips, L. J. 1986. Mental muddles. In M. Brand and R. M. Harnish, eds., *The representation of knowledge and belief*, 258 — 286. Tucson, Ariz. : University of Arizona Press.

Rips, L. J. 1994. *The psychology of proof. : Deductive reasoning in human thinking.* Cambridge, Mass. : MIT Press.

Rips, L. J., E. J. Shoben, and E. E. Smith. 1973. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior* 12, 120.

Rosch, E. B. 1973. On the internal structure of perceptual and semantic categories. In T. E. Moore, ed., *Cognitive development and the acquisition of language*, 111—144. New York: Academic Press.

Rosch, E. B., and C. B. Mervis. 1975. Family resemblances: Studies

in the internal structure of categories. *Cognitive Psychology* 7, 573—605.

Rosenbloom, P. S., J. E. Laird and A. Newell, eds. 1993. *The Soar papers: Research on integrated intelligence*. Cambridge, Mass.: MIT Press.

Rumelhart, D. E. 1980. Schemata: The building blocks of cognition. In R. Spiro, B. Bruce, and W. Brewer, eds., *Theoretical issues in reading comprehension*, 33—58. Hillsdale, N. J.: Erlbaum.

Rumelhart, D. E., and J. L. McClelland. 1982. An interactive activation model of context effects in letter perception. Part 2, The contextual enhancement effect and some tests and extensions of the model. *Psychological Review* 89, 60—94.

Rumelhart, D. E., J. L. McClelland, and the PDP Research Group. 1986. *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, Mass.: MIT Press.

Salomon, G., ed. 1993. *Distributed cognitions*. Cambridge: Cambridge University Press.

Schank, P., and M. Ranney. 1991. Modeling an experimental study of explanatory coherence. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, 892—897. Hillsdale, N. J.: Erlbaum.

Schank, P., and M. Ranney. 1992. Assessing explanatory coherence: A new method for integrating verbal data with models of on-line belief revision. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 599-604. Hillsdale, N. J.: Erlbaum.

Schank, R. C. 1982. *Dynamic memory: A theory of reminding and learning in computers and people*. New York: Cambridge University Press.

Schank, R. C. 1986. *Explanation patterns: Understanding mechanically and creatively*. Hillsdale, N. J.: Erlbaum.

Schank, R. C., and R. P. Abelson. 1977. *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, N. J.: Erlbaum.

Schank, R. C., A. Kass, and C. K. Riesbeck. 1994. *Inside case-based explanation*. Hillsdale, N. J.: Erlbaum.

Schmid, G. B. 1991. Chaos theory and schizophrenia: Elementary aspects. *Psychopathology* 24, 185 -198.

Schmitt F. F. ed. 1994. *Socializing epistemology*. Lanham. Md.: University Press of America.

Searle, J. 1992. *The rediscovery of the mind*. Cambridge, Mass.: MIT Press.

Seidenberg, M. S., and J. L. McClelland. 1989. A distributed, developmental model of word recognition and naming. *Psychological Review* 96, 523—568.

Shanon, B. 1993. *The representaional and the presentational*. New York: Harvester Wheatsheaf.

Shastri, L., and V. Ajjanagadde. 1993. From simple associations to systematic reasoning: A connectionist representation of rules, variables, and dynamic bindings. *Behavioral and Brain Sciences* 16, 417—494.

Shelley, C. P. In press. Visual abductive reasoning in archaeology. *Philosophy of Science*.

Shepard, R. N., and J. Metzler. 1971. Mental rotation of three-dimensional objects. *Science* 171, 701—703.

Skarda, C. A., and W. J. Freeman. 1987. How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences* 10, 161—195.

Smith, B. C. 1991. The owl and the electric encyclopedia. *Artificial Intelligence* 47, 251—288.

Smith, E. E. 1989. Concepts and induction. In M. I. Posner, ed., *Foundations of cognitive science*, 501 — 526. Cambridge, Mass.: MIT Press.

Smith, E. E., C. Langston, and R. Nisbett. 1992. The case for rules in reasoning. *Cognitive Science*, 16, 1—40.

Smith, E. E., and D. L. Medin. 1981. *Categories and concepts*. Cambridge, Mass.: Harvard University Press.

Smith, E. E., D. N. Osherson, L. J. Rips, and M. Keane. 1988.

Combining prototypes: A selective modification model. *Cognitive Science*. 12. 485—527.

Smith, L. B., and E. Thelen, eds. 1993. *A dynamic systems approach to development: Applications*. Cambridge, Mass: MIT Press.

Smolensky, P. 1990. Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence* 46, 159—217.

Solomon, M. 1994. Social empiricism. *Nous* 28, 325—343.

Spellman, B. A., and K. J. Holyoak. 1993. An inhibitory mechanism for goaldirected analogical mapping. In *Proceedings of the Fifteenth Annual confernce of the Cognitive Science Society*, 947—952. Hillsdale, N. J.: Erlbaum.

Stabler, E. P. 1992. *The logical approach to syntax*. Cambridge, Mass.: MIT Press.

Stillings, N. A., S. E. Weisler, C. H. Chase, M. H. Feinstein, J. L. Garfield, and E. L. Rissland. 1995. *Cognitive science: An introduction*. 2nd ed. Cambridge, Mass.: MIT Press.

St. John, M. F. 1992. The story gestalt: A model of knowledge-intensive processes in text comprehension. *Cognitive Science* 16, 271-306.

Suchman, L. 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge: Cambridge University Press.

Sun, R. 1994. *Integrating rules and connectionism for robust commonsense reasoning*. New York: Wiley.

Thagard, P. 1988. *Computational philosophy of science*. Cambridge, Mass.: MIT Press.

Thagard, P. 1989. Explanatory coherence. *Behavioral and Brian Sciences* 12, 435—467.

Thagard, P. 1992. *Conceptual revolutions*. Princeton, N. J.: Princeton University Press.

Thagard, P. 1993. Societies of minds: Science as distributed computing. *Studies in History and Philosophy of Science* 24, 49—67.

Thagard, P. 1994. Mind, society, and the growth of knowledge. *Philosophy of Science* 61, 629 –645.

Thagard, P., K. J. Holyoak, G. Nelson, and D. Gochfeld. 1990. Analog retrieval by constraint satisfaction. *Artificial Intelligence* 46, 259—310.

Thagard, P., and E. Millgram. 1995. Inference to the best plan: A coherence theory of decision. In A. Ram and D. B. Leake, eds. *Coal-driven learning*, 439-454. Cambridge, Mass.: MIT Press.

Thelen, E., and L. B. Smith. 1994. *A dynamic systems approach to the development of cognition and action.* Cambridge, Mass.: MIT Press.

Touretzky, D., and G. Hinton. 1988. A distributed production system. *Cognitive Science* 12, 423– 466.

Towell G. G., and J. W. Shavlik. 1994. Refining symbolic knowledge using neural networks. In R. Michalski and G. Tecuci, eds., *Machine learning: A multistrategy approach*. Vol. IV. 405-429. San Francisco: Morgan Kaufmann.

Tversky, A., and D. Kahneman. 1993. Extensional versus intensional reasoning: The conjunction fallacy in probability judgments. *Psychological Review* 90, 293—315.

Tye. M. 1991. *The imagery debate.* Cambridge, Mass.: MIT Press.

van Geert, P. 1991. A dynamic systems model of cognitive and language growth. *Psychological Review* 98, 3–53.

van Gelder, T. 1995. What might cognition be, if not computation? *Journal of Philosophy* 92, 345–381.

van Gelder, T., and R. Port. 1995. It's about time: An overview of the dynamical approach to cognition. In R. Port and T. van Gelder, eds., *Mind as motion: Explorations in the dynamics of cognition*, 1–43. Cambridge, Mass.: MIT Press.

von Eckardt, B. 1993. *What is cognitive science?* Cambridge, Mass.: MIT Press.

Waldrop, W. M. 1992. *Complexity: The emerging science at the edge*

of order and chaos. New York: Simon and Schuster.

Wason, P. C. 1966. Reasoning. In B. M. Foss, ed., *New horizons in psychology*. Harmondsworth: Penguin.

Watson, J. B. 1913. Psychology as the behaviorist views it. *Psychological Review* 20, 158—177.

Wharton, C. M., K. J. Holyoak, P. E. Downing, T. E. Lange, T. D. Wickens, and E. R. Melz. 1994. Below the surface: Analogical similarity and retrieval competition in reminding. *Cognitive Psychology* 26, 64—101.

Widrow, B., D. E. Rumelhart and M. A. Lehr. 1994. Neural networks: Applications in industry. *Communications of the ACM* 37(March), 93—105

Winograd, T., and F. Flores. 1986. *Understanding computers and cognition*. Reading, Mass.: Addison-Wesley.

Winston, P. 1993. *Artificial intelligence*. 3rd ed. Reading, Mass.: Addison-Wesley.

Wong, A. K. C., S. W. Lu, and M. Rioux. 1989. Recognition and shape synthesis of 3-D objects based on attributed hypergraphs. *IEEE Transactions of Pattern Analysis and Machine Intelligence* II(3), 279—289.